

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**STATISTICAL MODELING FOR DARK ENERGY AND
ASSOCIATED COSMOLOGICAL CONSTANTS**

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS AND APPLIED MATHEMATICS

by

Tracy Holsclaw

June 2011

The Dissertation of Tracy Holsclaw
is approved:

Professor Herbert Lee, Co-chair

Professor Bruno Sansó, Co-chair

Professor Athanasios Kottas

Dr. Katrin Heitmann

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Tracy Holsclaw

2011

Table of Contents

List of Figures	v
List of Tables	vi
Abstract	vii
Dedication	viii
Acknowledgments	ix
1 Introduction	1
1.1 What is in this thesis?	1
1.2 Cosmology and Dark Energy Developed	4
1.2.1 Type-Ia Supernovae	4
1.2.2 Dark Energy Model of the Universe	6
1.2.3 Bayesian Analysis and MCMC	9
1.2.4 Simulated Supernova Datasets	10
1.3 Conventional Parametric Models for the Dark Energy EOS	11
1.3.1 Model 1 - $w(z) = a$	14
1.3.2 Model 2 - $w(z) = a + bz$	16
1.3.3 Model 3 - $w(z) = a + b(\frac{1}{1+z} - 1)$	18
1.3.4 Discussion on Parameter Interdependencies	20
1.3.5 Remarks	22
2 Non-parametric Models for $w(z)$	24
2.1 Gaussian Process Models	26
2.1.1 Integration of a GP	26
2.1.2 General Example	28
2.2 Non-Parametric Models for the Dark Energy EOS	32
2.2.1 Model 4 - $w(z) \sim \text{GP}$	32
2.2.2 Model 5 - $w(z)$ is Modeled by a Damped Hermite Polynomial Basis	37
2.2.3 Conclusions	41
2.3 Hypothesis Testing and Model Comparison	42
2.3.1 Hypothesis Testing	42
2.3.2 Model Comparison	43
2.3.3 Conclusions	45
2.4 Real Datasets	47
2.4.1 Analysis and Results	49

2.4.2	Conclusions	54
3	Incorporating CMB and BAO Data	55
3.1	Simulated Data	57
3.2	Cosmic Microwave Background Data	58
3.3	Baryon Acoustic Oscillation Data	59
3.4	Results for Combined Data Sources	60
3.5	Real Data Analysis	64
4	Experimental Design	66
4.1	Cluster Analysis	66
4.2	Future Telescope Surveys	72
4.2.1	SNe Designed Experiments	75
4.2.2	Results	77
5	Conclusion	78
5.1	Future Work	80
A	GP Model: Details	82
A.1	Correlation Function Relationships	82
A.2	Integration of Correlation Functions	84
A.3	Localize Search Algorithm	85
A.4	GP Algorithm for the Inverse Method	86
A.5	MCMC Issues	87
A.6	Correlation Parameter	89
	Bibliography	91

List of Figures

1.1	Derivative Method for $w(z)$	7
1.2	Parameter Correlations	10
1.3	Simulated SNe Data	12
1.4	True Form of $w(z)$ for the Simulated SNe Data	12
1.5	Model 1 of $w(z)$	15
1.6	Model 2: $w(z)$	17
1.7	Model 3: $w(z)$	19
1.8	$w(z)$ Curves for Fixed Ω_m Values	21
2.1	Differentiated and Integrated GP Models	31
2.2	Model 4: $w(z)$	36
2.3	Damped Hermite Polynomials	39
2.4	Model 5: $w(z)$	40
2.5	Real SNe Data - Distance-Redshift Relation	48
2.6	Real SNe Data - Standard Error Associated with Distance Measures (μ)	48
2.7	Real Data - Model 1: $w(z)$	49
2.8	Real Data - Model 2: $w(z)$	52
2.9	Real Data - Model 3: $w(z)$	52
2.10	Real Data - Model 4: $w(z)$ is a GP ($\alpha = 1$)	52
2.11	Real Data - Model 4: $w(z)$ is a GP ($\alpha = 1.9999$)	53
2.12	Real Data - Model 5: $w(z)$	53
3.1	GP Model - SNe	60
3.2	GP Model - SNe+CMB	61
3.3	GP Model - SNe+BAO	61
3.4	GP Model - SNe+CMB+BAO	61
3.5	GP Model - BAO	62
3.6	GP Model - BAO+CMB	62
3.7	GP Models for SNe(Union 2)+BAO+CMB Data	65
4.1	Two-way Plots for Telescopes	68
4.2	Cluster Analysis Results	71
4.3	Redshift Distributions	73
A.1	Correlation Parameter Priors	89
A.2	Model 4: $w(z)$	91
A.3	Prior and Posterior for ρ	91

List of Tables

1.1	Model 1 - Posterior 95% PIs	14
1.2	Model 2 - Posterior 95% PIs	17
1.3	Model 3 - Posterior 95% PIs	19
2.1	Model 4 - Posterior 95% PIs	35
2.2	Model 5 - Posterior 95% PIs	40
2.3	Model Comparison - Fixed Δ and Ω_m	45
2.4	Model Comparison - Variable Δ and Ω_m	46
2.5	All Models - 95% Posterior PIs for Real SNe Data	50
2.6	Model Comparison for the Real Data	51
3.1	Model 4 - Simulated Data - Posterior 95% PIs	63
3.2	Model 4 - Union 2 SNe Dataset - Posterior 95% PIs	64
4.1	Cluster Analysis	70
4.2	Telescope Missions	73
4.3	Telescope Mission Comparison	77

Abstract

Statistical Modeling for Dark Energy and Associated Cosmological Constants

by

Tracy Holsclaw

Our endeavor has been to answer one of science's important questions, mainly about the nature of dark energy. We have worked alongside cosmologists to better understand our Universe and in due course have developed some useful statistical methods for undertaking analysis of derivative processes, model selection, and experimental design. There are no direct measures available for the posited dark energy; its form must be inferred from other data sources like supernova, cosmic microwave background radiation, or baryon acoustic oscillation. The dark energy equation of state is a second derivative process embedded in a non-linear transform when related to the observable data. An inverse method is required to coherently model the dark energy equation of state and relate its fit back to the observed data, which requires two integrations. In general, parametric forms have been used to model the dark energy equation of state because of the complexity of the inverse problem. We show the form of dark energy can be modeled with a non-parametric Gaussian process which can be integrated by properties of the stochastic process. This results in a computationally efficient algorithm for the integrations. This inverse statistical method of estimating functions of derivatives with Gaussian processes is generalizable to many other applications. Additionally, we show the benefits of this modeling for the dark energy equation of state through model comparison methods that can handle both parametric and non-parametric models. Finally, we compare future data collection missions using the Gaussian process model in an experimental design setting.

To Gregory,
my loving and very patient husband,
as I had many tired days and long hours of studying.

To my family,
yes all of them,
for their continual support.

Acknowledgments

I am grateful for the two statisticians that lead me through this entire process and project. Herbie Lee and Bruno Sansó met week after week to discuss the project and guide the statistical development. In the beginning, there was no assurance that there would be a satisfactory solution to this problem. The first six months yielded poor results but they reshaped and refocused the project without being discouraged. They allowed me the freedom to pursue the research at my pace and continued to support me. They were quite patient especially when things took longer than expected, I had yet another bug in the code, or I was just having a bad week with the project. They always were there to encourage me to keep at it. David Higdon provided input and oversight on the project and was part of the initial arrangement for the collaboration with Los Alamos National Laboratory.

A very big thanks to the cosmologists Katrin Heitmann, Ujjaini Alam, and Salman Habib for the inspiration for this project. The exploration of the dark energy equation of state is truly a joint effort from both the cosmology and statistical teams. I am grateful for the endless hours, many exchanges, and detailed explanations from the cosmologists about their field of study and their willingness to start at a very rudimentary level because this was a new field of scientific work for me. I thank Katrin Heitmann for all of the correspondence in regard to the joint papers. I was able to learn so much from her excellent writing and research skills. I extend deep appreciation to Ujjaini Alam for her patience in explaining all of the “simple” cosmological things from day one and the many data files she produced. And to Salman Habib for his detailed review of the papers and insight into the cosmology that was invaluable in shaping the project and direction.

Also, thanks to the Los Alamos National Laboratory staff for keeping track of the project as it progressed; along with the ISSDM staff at UCSC for arranging the teleconferences, quarterly collaboration meetings, and the yearly ISSDM conference. And for those

involved with all of the travel arrangements for our two teams over the course of the four years.

I want to thank my professors who taught courses and imparted knowledge that was then applied to this project. Specifically, I want to thank, Herbie Lee for teaching Bayesian Methods I: introducing the Bayesian framework and Markov chain Monte Carlo algorithm (ensuring for his sake the c in chain is lower case and data is used as a plural in this entire document). To Bruno Sansó for his courses on spatial statistics (I will never forget the Matérn correlation function). To Raquel Prado for the wonderful and foundational Bayesian modeling course. To Athanasios Kottas for his theoretic courses on stochastic processes, probability theory, and Bayesian non-parametrics. I am very appreciative for his willingness to serve on my committees and all of his encouraging words along the way.

I am so very thankful for the other great graduate students I had a privilege of studying alongside and preparing for the first year exams and other hardships of graduate life. I hope they have all enjoyed this experience as much as I have. I am especially am thankful to those I could ask the basic questions when I forgot something, so I would not have to bother my advisors yet again.

And to my husband who did not think it good enough to have the dissertation dedicated to him but also wanted to be in the acknowledgments. Thank you for driving me on those days I was just too tired or sick to make it. And for being there on days when I hit those walls of frustration.

Chapter 1

Introduction

1.1 What is in this thesis?

The Universe is expanding at an accelerating rate, but what is causing this phenomenon? Some have posited the existence of a mysterious entity coined dark energy, yet no one has thus far proposed a way of measuring or even proving that it exists. There is a good possibility that it does exist, in which case we desire to make inference as to the physical and mathematical nature of this inexplicable phenomenon. Specifically, we want to make inference on the equation of state of this possible force, dark energy, in so much as to determine its form and account for any historical changes in its nature.

There are several probing questions pushing the forefront of discovery in cosmology; this is one of them. It requires new methods and statistical applications to answer this question. In no way is the current cosmology straightforward or the equations simple to analyze. In Chapter 1, we lay the groundwork of the current non-linear cosmology model and how applied statistics is currently being used to address the questions about the dark energy equation of state.

Currently, this unknown driver of our accelerating Universe, dark energy, can only

be studied through its pressure-density relation by means of auxiliary observable data relations like the distance-redshift relation of supernovae. Cosmologists wonder if the equation of state for dark energy ($w(z)$) could be as simple as an inverse relationship that has been constant since the beginning of the Universe (mathematically does $w(z) = -1$?). In Chapter 2, we answer the important question: is $w(z) = -1$ or at least prepare a framework for this question to be answered in the future as supernova data quality and quantity is improved.

Checking this hypothesis is no small feat, as the equation of state (EOS) for dark energy is a second derivative process of the data and is embedded in a non-linear cosmology equation relating distance and redshift. Scientists rightly address this as an inverse model, this changes the derivatives into integrations over the redshift domain. $w(z)$ can only be fit by a one or two parameter model under this framework. Alternatively, we propose a non-parametric Gaussian process (GP) model that allows a flexible fit of the equation of state of dark energy. This GP model is shown to have nearly the same number of effective parameters as the two parameter model, while obtaining tighter estimates of the equation of state for dark energy ($w(z)$) and is more flexible in its fit than any other current model. The statistical accomplishment is not only that we have attempted to fit a non-parametric GP model to a very difficult cosmology application but that we have an innovative way of solving an inverse problem exploiting the integrable properties of the GP. We show how the GP can give a flexible fit with more constrained probability intervals (PIs) to help answer if $w(z) = -1$.

Cosmologists want to know what data is needed to better constrain the current fit for $w(z) = -1$. There may not be sufficient supernova data to support the framework for answering such a simple question. In Chapter 3, we examine how baryon acoustic oscillation (BAO) and cosmic microwave background (CMB) data can help constrain the equation of state of dark energy ($w(z)$) and the matter density constant of the Universe (Ω_m). Other models have not been sufficiently capable of handling the high redshift CMB data in a

coherent manner or the newer BAO measures that require integrations at high redshift. We believe that the GP model we present is coherent for both local and distant/historical measurements that are needed to better constrain the equation of state of dark energy.

The form of $w(z)$ can be further constrained by improving the quantity and quality of the supernova data. While much of the work on retrieving better quality data is solely in the hands of the astronomer, statistically we should be able to answer a foundational question that bridges the theoretical cosmologist and observational astronomer about where more data could be needed. Chapter 4 addresses this issue of where this data should be collected on the z (redshift) axis to best constrain $w(z)$ by comparing three possible telescope missions. Uncertainty may be lessened by collecting expensive observations in these future studies. This is not only important to the cosmologist and astronomer, but also to the decision maker who must allocate resources for new telescopes or purchase time on existing telescopes.

Our endeavor has been to answer one of science's important questions, mainly about the nature of dark energy. We have worked alongside cosmologists to better understand our Universe and in the meantime have developed some useful statistical methods for undertaking analysis of a derivative process (like the dark energy equation of state) where no direct measurements are available. In the process of tackling the cosmological problem, we have developed statistical methodology to estimate the derivative of a curve that is of general interest. Our GP model requires the non-parametric fit of interest to be integrated, which we show can be done by properties of the stochastic process. Here we develop and employ innovative statistical methods for fitting derivative processes to give a non-parametric form for the dark energy equation of state and reduce uncertainty around that estimate.

1.2 Cosmology and Dark Energy Developed

It has been assumed since the 1920's that the Universe is expanding (de Sitter, 1934). Surprisingly in 1998, observational evidence started accumulating that favor a model of the Universe that is expanding at an accelerating rate (Perlmutter et al., 1999). An unknown entity termed dark energy (DE) is one possible explanation for this acceleration (Riess et al., 1998). Currently, the nature of dark energy can be investigated by studying its equation of state (EOS), that is the relationship of its pressure to its density. The hypothesized dark energy is not directly detectable or measurable, so other means of investigation are employed to learn about this possible mysterious influence.

To explore the theory of dark energy, observations are needed to verify the current cosmological models and estimate the unknown parameters of interest. We investigate some of these probes to gain a better understanding of the nature of this posited dark energy. These probes provide observations that have distance-redshift relations, which in turn provides information about the cosmological parameters and the dark energy EOS. The abundance of galaxy clusters, a measure of the background photons passing through hot clusters called the integrated Sachs-Wolfe effect, weak lensing, type Ia supernovae (SNe), cosmic microwave background radiation (CMB), baryon acoustic oscillation (BAO) measured from the distribution of galaxies, and assumptions on the structure of the Universe can be used as measurable quantities to make inference as to the properties of dark energy (Genovese et al., 2009). We specifically focus on the best of these observations for making inference on the dark energy EOS.

1.2.1 Type-Ia Supernovae

One of the best measures of the expanding Universe is the abundant type-Ia supernovae because of their property as standardizable candles (Perlmutter et al., 1997; Lei-

Leibundgut, 2004; Riess et al., 1996b; Wood-Vasey et al., 2008). This property provides a distance measure and can be directly related to the measurement of the supernova's redshift (z) (Way et al., 2009). This distance-redshift relationship of supernovae and other probes are key in determining the value of the dimensionless matter density parameter for the Universe (Ω_m) and the dark energy EOS ($w(z)$) (Genovese et al., 2009; Huterer and Turner, 2001).

The peak luminosity (or luminosity distance D_L) when a supernova explodes provides information on its distance (Way et al., 2009). Over the course of a few weeks of observing a supernova explosion, the emitted light slowly grows to peak intensity and then fades. The peak luminosity is of greatest interest as it bears the distance information of the object; it can be obtained from fitting a light curve to the weeks of observations (Pskovskii, 1977). The brightness of the supernova is closely related to the shape of its light curve (Leibundgut, 2001). Nearby supernova data are used to assist in estimating the color, light curve shape, and peak luminosity in a complex light curve fitting process (Wood-Vasey et al., 2007; Riess et al., 1996a). In this discussion, we assume that the previous work of the astronomer and the light curve fitting protocol they have developed is correct and provides valid peak luminosity values ($D_L = 10^{\frac{\mu-25}{5}}$) and error bars for these measurements (τ^2). Most of the variability in distance measurements is assumed to be due to differences in the supernovae at low and high redshift and needed color correcting for reddening from intergalactic dust. This includes complex astronomy, filtering, multiple measurements per supernovae, classification of the supernova as a type-Ia, and light curve fitting processes (Guy et al., 2005).

Redshift (z) is the other measurable quantity of a supernova and it results from power spectrum analysis (Dodelson, 2003). We are referring to the cosmological redshift of an object, which is not quite the same as the typical redshift; this type of cosmological redshift encompasses the stretching of space between objects (Filippenko, 1997). The redshift

and distance measures are related through an equation which contains the unknown dark energy EOS and some of the cosmological parameters (Way et al., 2009). We examine this distance-redshift relation to shed light on the acceleration of the Universe and about the mysterious dark energy.

1.2.2 Dark Energy Model of the Universe

To construct the posited dark energy model of the Universe, many cosmological assumptions are needed and are based on the work of general relativity with the Friedmann-Robertson-Walker metric and other assumptions of the Universe. Other foundational work includes estimation of the Hubble parameter and the speed of light. We assume to know the speed of light (c) precisely because it has been measured quite accurately. But we need to estimate other unknown parameters like the Hubble parameter (H_0), an offset parameter due to uncertainty in the supernovae measurement process (M), and the matter density of the Universe (Ω_m) in the current mathematical representation of the Universe. We add a radiation term to the cosmological model ($\Omega_r(s)(1+s)^4$) for completeness but it is unnecessary when analyzing current low-redshift supernovae observations. Typically, we assume that $\Omega_r(z)$ is zero when working with just low redshift data like SNe. This term only becomes important when high redshift data or equations are being analyzed like in the case of CMB and BAO data.

Finally, we come to the distance-redshift relation for SNe data accounting for the comoving distance ($r(z)$) to the object (Huterer and Turner, 1999; Genovese et al., 2009). Equations (1.1) and (1.2) are non-linear equations that relate the distance and redshift for SNe.

$$r(z) = \frac{1}{c(1+z)} 10^{\frac{\mu-25-M}{5}} \quad (1.1)$$

$$r(z) = \frac{1}{H_0} \int_0^z \left(\Omega_r(s)(1+s)^4 + \Omega_m(1+s)^3 + (1 - \Omega_r(s) - \Omega_m)(1+s)^3 e^{-3 \int_0^s \frac{-w(u)}{1+u} du} \right)^{-1/2} ds \quad (1.2)$$

The dark energy EOS ($w(z)$), a second derivative process of the fit, is the function of main interest. There are several other unknown parameters that appear in this distance-

redshift equation: Ω_m and H_0 . There are two quite different approaches to estimate the unknown function $w(z)$ and these unknown parameters. One approach is to fit equation (1.2) directly to the SNe data then find one or both derivatives, r' or r'' respectively. The data are discrete and noisy so it is impossible to differentiate directly without using some sort of smoothing. The second derivative of such curve would be directly related to $w(z)$ through a the non-linear transform (Sahni and Starobinsky, 2006; Saini et al., 2000; Weller and Albrecht, 2002). This approach is no longer favored because it loses information through the differentiation mechanism. Additionally, the fit of the second derivative process estimates low redshift values poorly and gives large uncertainty bands for high z values; Weller and Albrecht (2002) shows these fits of $w(z)$ in their Figure 6. We include a Figure 1.1 to show the result of fitting the SNe data directly with non-linear regression ($\log(\mu) = \beta_0 + \beta_1 \log(z) + \beta_2 (\log(z))^2$) and then differentiating to obtain $w(z)$ for a simulated dataset where the truth for $w(z)$ is known to be a constant negative one (dashed red line). The estimated fit for $w(z)$ in this plot is poor, actually it does not fit the physics of $w(z)$, as seen in the behavior around $z=0$ and for z greater than one, as well as the PI bands being quite wide.

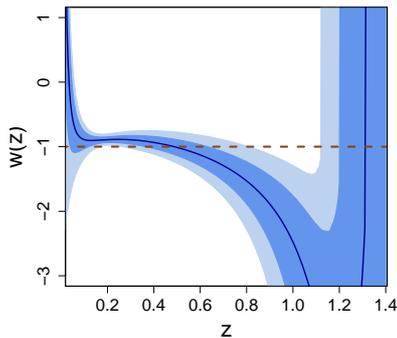


Figure 1.1: z vs. $w(z)$ for simulated SNe data with $n=557$ observations. Mean (dark blue), 68% PI (blue), 95% PI (light blue), and truth (red dashed line)

In comparison to using this differentiation approach, some authors now view this

as an inverse problem and a priori assume a parametric form for $w(z)$ (Huterer and Turner, 2001; Astier, 2000). For most parametric forms of $w(z)$, the inner integral is analytic, the outer integration can be done numerically, and the parameters can be evaluated in a Markov chain Monte Carlo (MCMC) algorithm (Gamerman and Lopes, 2006). In the algorithm for the inverse method, one assumes/guesses a form for $w(z)$, not for the data; the proposed fit of $w(z)$ is integrated and transformed to check it against the data (Genovese et al., 2009). This provides inference for all of the parameters of interest.

We exclusively view this as an inverse problem where we assume a form for $w(z)$ and use SNe data. The SNe data is in terms of z , μ , and τ , where $\mu = 5 \log_{10}(c(1+z)r(z)) + 25$. Let $\mu_i = \alpha(z_i) + \epsilon_i$ where τ_i is approximately one standard deviation of uncertainty in μ_i and σ^2 is the estimated variance of the data. We have Normal distributed errors: $\epsilon_i \sim N(0, \tau^2 \sigma^2)$ or $\mu_i \sim N(T(z), \tau^2 \sigma^2)$ where $T(z)$ is given in equations (1.3) and (1.4).

$$T(z) = 25 + M - 5 \log_{10}(H_0) + 5 \log_{10} \left(c(1+z) \int_0^z G(s) ds \right) \quad (1.3)$$

$$G(s) = \left(\Omega_r(s)(1+s)^4 + \Omega_m(1+s)^3 + (1 - \Omega_r(s) - \Omega_m)(1+s)^3 e^{-3 \int_0^s \frac{w(u)}{1+u} du} \right)^{-1/2} \quad (1.4)$$

Unfortunately, SNe data alone does not supply information about Hubble's constant (H_0). Both H_0 and M are unknown parameters but are indistinguishable from one another. A new parameter Δ is used to encompass the joint uncertainty of these two parameters. We fix the value of H_0 to an estimate (H_0^*) and then allow the uncertainty of that parameter to propagate into Δ . This leads to a new representation of the distance-redshift transform seen in equation (1.5). The likelihood is set up as follows for the SNe data: $L \propto \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2} \sum \left(\frac{\mu_i - T(z)}{\tau_i \sigma}\right)^2}$.

$$T(z) = 25 + \Delta - 5 \log_{10}(H_0^*) + 5 \log_{10} \left(c(1+z) \int_0^z G(s) ds \right) \quad (1.5)$$

1.2.3 Bayesian Analysis and MCMC

We have shown that dark energy EOS ($w(z)$) can be estimated by using the distance-redshift relationship of SNe data. The distance-redshift relation is a non-linear transform that ultimately is embedded in a likelihood equation. At this point, we have a choice on what type of statistical methodology best suits this problem. There is no simple closed form solution to such a problem because of the non-linearity and complexity involved in the likelihood; some type of numerical algorithm is necessary. Much of the cosmology literature uses a Bayesian approach to this problem and open source code based on Markov chain Monte Carlo (MCMC) algorithm is available for some of the simpler parametric models of $w(z)$. The Bayesian approach also lends itself here because we have prior information for Ω_m and that type of prior information is not easily incorporated into other statistical frameworks.

Auxiliary Variables Ω_m , Δ , and σ^2

A Bayesian approach assumes all unknown parameters to follow a distribution (Gelman et al., 2004). We have several parameters of interest: Ω_m , Δ , σ^2 , and any parameters that comprise $w(z)$ depending on the model. Both Δ and σ^2 posterior full conditional distributions can be sampled via an MCMC algorithm as Gibbs steps.

We found that Ω_m and Δ along with the parameters that comprise $w(z)$ (when $w(z)$ has a parametric form) are correlated to one another. Figure 1.2 shows two-way plots of unknown parameters for simulated data with $w(z) = -1$. The model being fit is $w(z) = a$ with unknown Δ and Ω_m . The three parameters are sampled in a single Metropolis-Hastings step of the algorithm using a joint proposal. The multivariate Normal joint proposal requires a covariance matrix to be tuned, which can require a few thousand initializing runs. This type of joint proposal is necessary for stable uncorrelated Markov chains (see Appendix A.5 for a further discussion on MCMC issues).

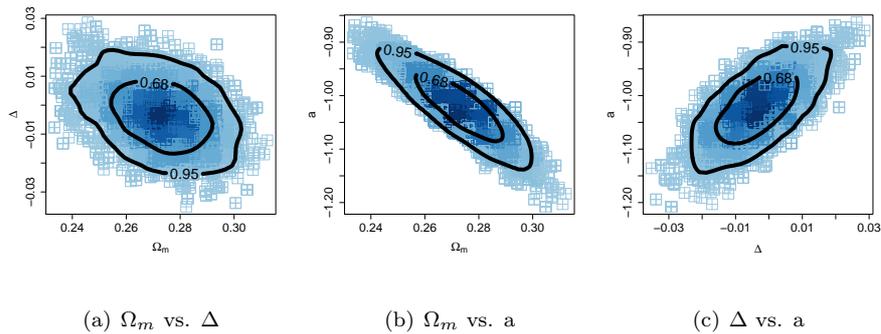


Figure 1.2: Two-way plots for Ω_m , Δ , and the parameter in $w(z)$ when $w(z) = a$

Because Ω_m and Δ are parameters and we consider this a Bayesian framework; they both need priors. The prior for Δ is a flat Uniform. Whereas, Ω_m has an informative prior: $\pi(\Omega_m) \sim N(0.27, 0.04^2)$ which is based on other data sources like BAO and CMB, which are independent of SNe data (Komatsu et al., 2010). In Chapter 3, BAO and CMB data are incorporated and then a flat Uniform prior is used over the entire range of possible values: $\pi(\Omega_m) \sim U(0, 1)$.

We use a non-informative prior for σ^2 and incorporate this into the Gibbs step: $\pi(\sigma^2) \propto \sigma^{-2}$. Because the error bars (τ_i) are accounting for the variation in the data, the true σ^2 should be approximately one. The conditional posterior distribution for σ^2 is shown in equation (1.6), where $T(z)$ is the transform given in equation (1.3):

$$\sigma^2|z, H_0, \Omega_m, M, w_0 \sim IG\left(\frac{n}{2}, \frac{1}{2} \sum_{i=1}^n \left(\frac{\mu_i - T(z_i)}{\tau_i}\right)^2\right) \quad (1.6)$$

1.2.4 Simulated Supernova Datasets

To model the distance-redshift relationship of SNe Type Ia, we need data. We could use real SNe data but simulated datasets have some benefits, the truth is known and we can test the models and methods for accuracy. This simulated SNe data is similar to what is expected to be obtained in the near future in the Joint Dark Energy Mission (JDEM)

(Aldering, 2005). Each supernova observed results in a redshift measurement (z), a peak luminosity used as a distance measure (μ) and its associated standard deviation (τ). The standard error (τ) is set at a constant 0.13 for every observation. These datasets reflect the expectation of collecting $n = 2023$ supernova observations. The true values of the parameters are fixed when generating the data at: $H_0 = 72$, $\Delta = 0$, and $\Omega_m = 0.27$. Therefore, when using the simulated data we can either assume to know the auxiliary variables Ω_m and Δ and solely focus on modeling $w(z)$ or we can let them vary.

We consider three simulated datasets to test the robustness of our analysis before applying them to real data. Figure 1.3 graphs the three datasets; they are called μ_1 , μ_2 , and μ_3 . The major difference in these datasets is the underlying form of $w(z)$, which can be seen in Figure 1.4. μ_1 has the simple truth being $w(z) = -1$ (Figure 1.4(a)). μ_2 has a $w(z)$ with slight curvature (Figure 1.4(b)) which should be fit well by one of the two-parameter models we plan on fitting (Linder, 2003; Chevallier and Polarski, 2001). Figure 1.4(c) shows μ_3 , which is a more complex form of $w(z)$ with a much faster rate of change at mid-redshift values. This form is already excluded based on the real data but is a good test case because it shows how the parametric forms can fail. The data values for z versus μ seen in Figure 1.3 values are very similar for all three of these models. In fact, they are nearly indistinguishable, which leads us to conclude that very different $w(z)$ curves (non-linearly transformed second derivative process of the data) produce nearly identical $\mu(z)$ data curves.

1.3 Conventional Parametric Models for the Dark Energy EOS

Many others are working to answer the question: is $w(z) = -1$? As background, we begin by reproducing some of the most common parametric models for the dark energy

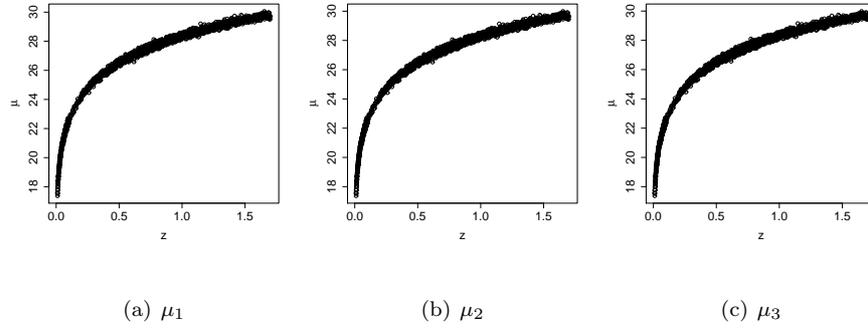


Figure 1.3: Graphs of the distance-redshift relation for the SNe data: z vs. μ_B

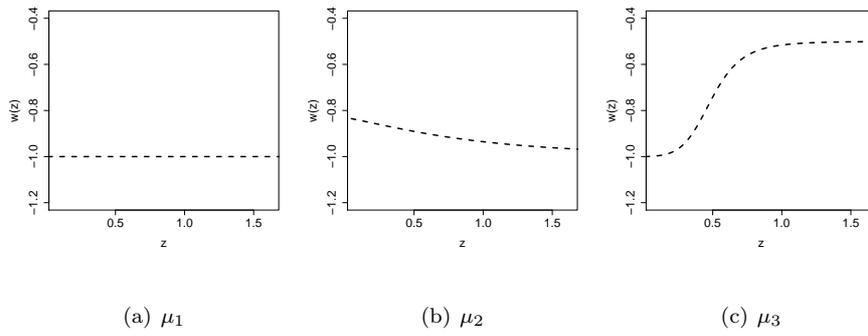


Figure 1.4: True $w(z)$ curves

EOS ($w(z)$). Our analysis includes Bayesian estimation for three of the most popular ansatz models for $w(z)$: $w(z) = a$, $w(z) = a+bz$, and $w(z) = a+b\left(\frac{1}{1+z} - 1\right)$. These models include the constant case, one linear and one non-linear parametric form for $w(z)$. We use similar priors for all of these models. We performed extensive prior sensitivity analysis that showed that rather non-informative priors for the parameters in $w(z)$ were appropriate. We use: $\pi(a) \sim U(-25, 1)$ and $\pi(b) \sim U(-25, 25)$. These rather standard models for this problem have been proposed and explored in previous work by Linder (2007) and many others.

Because we use simulated data (called datasets μ_1 , μ_2 , and μ_3), the truth for the models is known. The analysis for the coefficients, a and b and plots of $w(z)$ can be easily compared against the truth. For simulated dataset μ_1 , where $w(z) = -1$, the truth is simple with the first $w(z) = a$ model being the correct model. This means that for dataset μ_1 the underlying truth is $a = -1$ and $b = 0$ for these three models. Simulated dataset μ_2 , comes from a curve similar to the third model with parameters $a = -0.818$ and $b = 0.232$. Dataset μ_3 is a more complex function of $w(z)$ with quite a bit of curvature and does not correspond to any of these parametric models. The $w(z) = a$ model is not able to fit the data well but should give something like an average of $a = -0.619$.

For each parametric model, we show results for these three datasets. First, we fix Ω_m and Δ to the true values of the simulated data and solely explore the fit of $w(z)$. Second, we perform the analysis with Ω_m and Δ as unknown parameters that are estimated. Because we are using a Bayesian model, this analysis results in posterior distributions for these parameters; we display the 95% probability intervals (PIs) for all estimated parameters. The simulations are all run 10,000 times with an MCMC algorithm and all acceptance rates for the posteriors were around 20% and typically within 10-40%. The mixing of the Metropolis algorithm is acceptable in all cases. We show a mean fit for $w(z)$ as a solid line, the 68% probability interval in dark blue, the 95% probability interval in light blue, and the truth as a dashed line.

1.3.1 Model 1 - $w(z) = a$

The first model of interest is the simplest; we assume that $w(z)$ is constant and unchanged from the beginning of the Universe. This model is the easiest to work with and is best suited for traditional hypothesis testing as it results in a single value for $w(z)$ with a probability interval. However, the downside to this model is that it is incredibly rigid and does not allow for $w(z)$ to evolve through the history of the Universe and cannot indicate if $w(z)$ is curved. Many are interested in determining whether $w(z)$ is a cosmological constant: whether it be -1, -2/3, -1/3, or something else (Genovese et al., 2009). If it could be shown with great certainty that $w(z)$ is constant then this model would be sufficient, however currently there is too much uncertainty. In this model, we have $w(z) = a$ and this leads to a simplified form for $G(z)$ seen in equation (1.7). Table 1.1 and Figure 1.5 contain the results of this analysis.

$$G(z) = (\Omega_r(1+s)^4 + \Omega_m(1+s)^3 + (1 - \Omega_m - \Omega_r)(1+s)^3(1+s)^{3a})^{-1/2} \quad (1.7)$$

Table 1.1: Model 1 - Posterior 95% PIs

Dataset	a	Ω_m	Δ	σ^2
μ_1	(-1.016,-0.990)	0.27	0	(0.92,1.03)
μ_2	(-0.874,-0.851)	0.27	0	(0.92,1.03)
μ_3	(-0.928,-0.903)	0.27	0	(0.93,1.05)
μ_1	(-1.120,-0.923)	(0.246,0.296)	(-0.020, 0.015)	(0.92,1.03)
μ_2	(-0.940,-0.752)	(0.224,0.290)	(-0.020, 0.011)	(0.92,1.03)
μ_3	(-1.306,-1.069)	(0.325,0.366)	(-0.023, 0.011)	(0.92,1.03)

Analysis

We do not gain information about Ω_m from this analysis; as the prior and posterior are nearly identical. Instead Ω_m is included in the model to show the uncertainty it adds to the estimation of $w(z)$ (there are wider bands in the second row of graphs). Additionally, when Ω_m is added to the third dataset, $w(z)$ is poorly estimated. Figure 1.5(f)

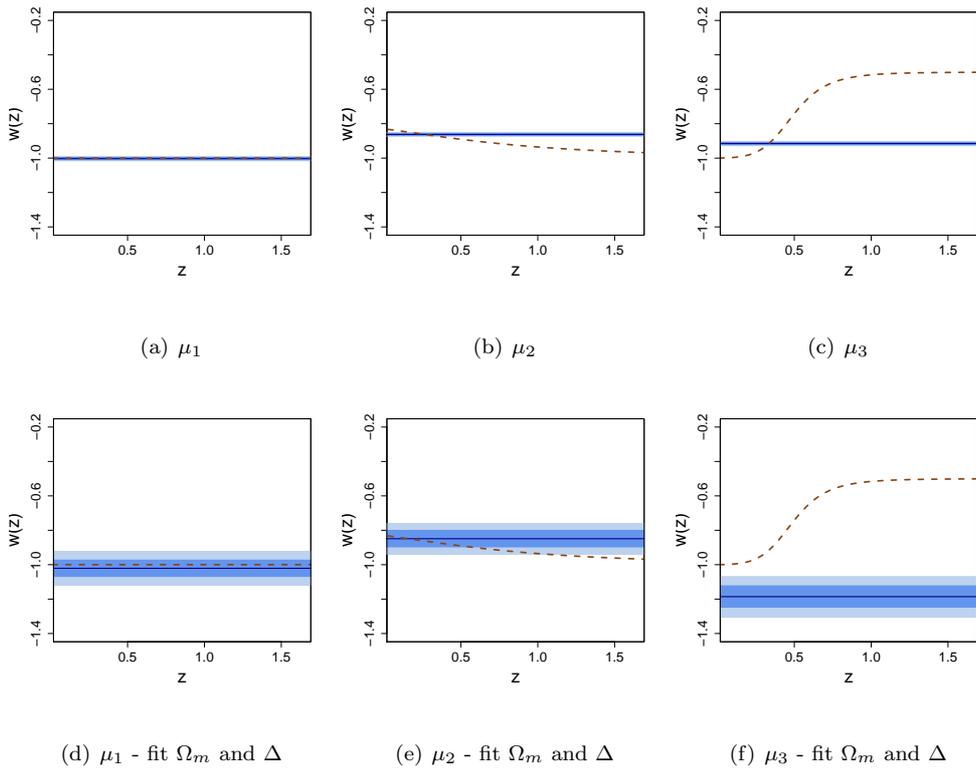


Figure 1.5: Model 1 fits of $w(z)$ for the three datasets μ_1 , μ_2 , and μ_3 . Mean (dark blue), 68% PI (blue), 95% PI (light blue), and truth (red dashed line)

has an especially poor fit for both Ω_m and $w(z)$; the fit for $w(z)$ does not touch the curve, which is completely counter-intuitive. This chapter concludes with a section on parameter interdependencies that should help explain this poor fit in the case when Ω_m is not well constrained.

We see in Table 1.1 and plots 1.5(a) and 1.5(d) that this model fits dataset μ_1 quite well. This is expected as the truth for dataset μ_1 is a constant. However, Model 1 ($w(z) = a$) does not fit datasets μ_2 or μ_3 well because the truth for $w(z)$ in these datasets is curved. Overall, Model 1 provides tight PI bands but it cannot make inference as to the shape of $w(z)$, as it inherently assumes a flat dark energy EOS.

1.3.2 Model 2 - $w(z) = a + bz$

Model 2 is a simple linear model for $w(z)$ with two coefficients and is explored in Cooray and Huterer (1999); Maor et al. (2001); Weller and Albrecht (2001). This model allows for a constant linear change over the history of the Universe for the dark energy EOS. Therefore, the assumption here is more relaxed than in Model 1, as it picks up a type of redshift dependency. We let $w(z) = a + bz$ which leads to a simplified version of $G(z)$:

$$G(z) = \left(\Omega_r(1+s)^4 + \Omega_m(1+s)^3 + (1 - \Omega_m - \Omega_r)(1+s)^{3(a-b+1)} e^{3bs} \right)^{-1/2}$$

The general results can be seen in Table 1.2. Graphical fits for the dark energy equation of state, $w(z)$, are in Figure 1.6 with a mean line in black, 68% probability bands in dark blue, and 95% probability bands in light blue. The true curve is displayed as a red dashed line.

Analysis

Model 2 does an adequate job of fitting datasets μ_1 and μ_2 ; the flat dataset and the slightly curved one. However, Model 2 does not capture the true $w(z)$ for μ_3 , as this

Table 1.2: Model 2 - Posterior 95% PIs

Dataset	a	b	Ω_m	Δ	σ^2
μ_1	(-1.048,-0.952)	(-0.207,0.177)	0.27	0	(0.91,1.03)
μ_2	(-0.880,-0.787)	(-0.298,0.069)	0.27	0	(0.92,1.03)
μ_3	(-1.110,-1.016)	(0.416,0.743)	0.27	0	(0.91,1.03)
μ_1	(-1.114,-0.913)	(-1.156,0.408)	(0.205,0.335)	(-0.022,0.017)	(0.92,1.03)
μ_2	(-0.941,-0.763)	(-1.130,0.214)	(0.217,0.348)	(-0.020,0.016)	(0.92,1.03)
μ_3	(-1.278,-0.973)	(-0.505,0.780)	(0.223,0.377)	(-0.024,0.012)	(0.92,1.03)

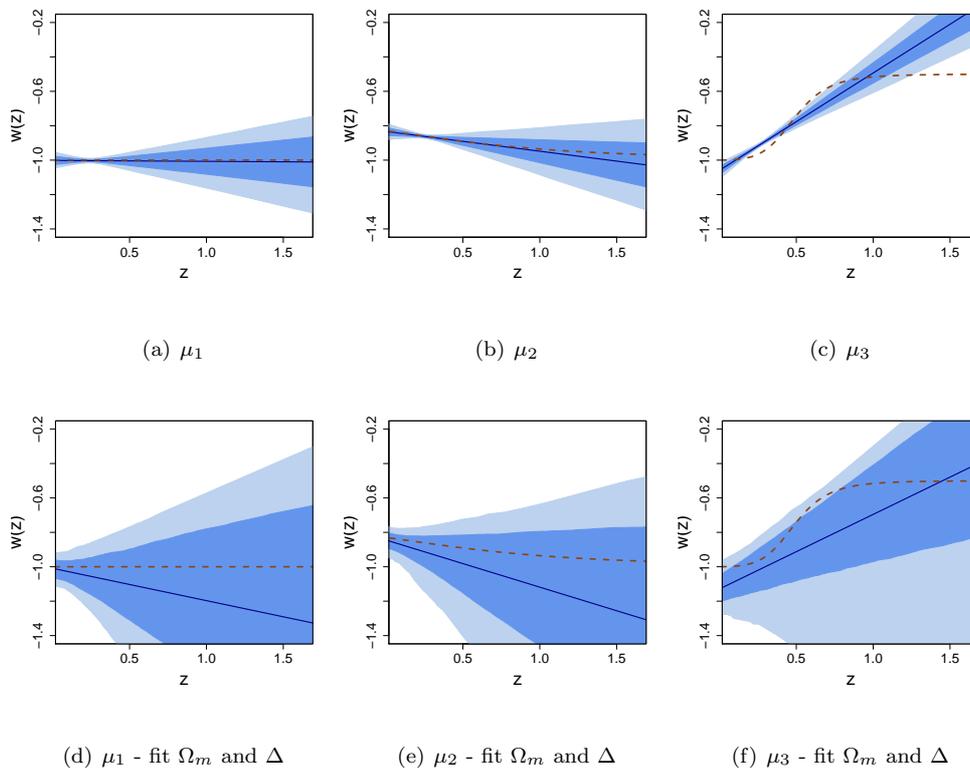


Figure 1.6: Model 2 fits of $w(z)$ for the three datasets μ_1 , μ_2 , and μ_3 . Mean (dark blue), 68% PI (blue), 95% PI (light blue), and truth (red dashed line)

dataset does not have a linear form. This is apparent in Figure 1.6; the dashed true line for $w(z)$ is outside the probability bands.

This linear model estimates the value for Ω_m and Δ fairly well in all cases, which is an improvement from Model 1. When Ω_m and Δ are added to the model the probability bands are much wider resulting in more uncertainty in the mean fit of $w(z)$. According to this analysis the parameters a and b are more correlated with each other when Ω_m and Δ are fixed. Overall, this simple linear model provides a basic way to examine the possibility of a change of the dark energy EOS over the redshift range but it only allows for a rigid linear change.

1.3.3 Model 3 - $w(z) = a + b(\frac{1}{1+z} - 1)$

This two parameter model has been recommended by some of the cosmological literature as a form of interest for $w(z)$. Linder has been one of the main proponents of this model as a robust form for an equation of state (EOS) with monotonic behavior (Linder, 2003). Linder advocates using the parameterization: $w(z) = a^* + b^*(1 - \frac{1}{1+z})$ to avoid a z dependence issues (Linder, 2006). This two parameter model assumes a form of $w(z)$ that is somewhat more flexible than a linear model but has its limitations as to what forms of the dark energy EOS it can model. Linder examines higher order polynomial fits in subsequent literature and find they do not estimate $w(z)$ correctly and oscillate heavily and must be truncated for high z values (Linder, 2007). Here we let $w(z) = a + b(\frac{1}{1+z} - 1) = a + \frac{-bz}{1+z}$ in the $r(z)$ equation and do not include any higher order terms. For this model there is a distance-redshift relation of the form:

$$G(z) = \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^{3(a-b+1)} e^{\frac{3bs}{1+s}} \right)^{-1/2}.$$

All of the unknown parameters were sampled jointly because they are correlated to one another. The results of this analysis can be seen in Tables 1.3 and Figure 1.7.

Table 1.3: Model 3 - Posterior 95% PIs

Dataset	a	b	Ω_m	Δ	σ^2
μ_1	(-1.068,-0.935)	(-0.358,0.393)	0.27	0	(0.91,1.03)
μ_2	(-0.886,-0.766)	(-0.122,0.534)	0.27	0	(0.92,1.03)
μ_3	(-1.164,-1.049)	(-1.363,-0.757)	0.27	0	(0.92,1.03)
μ_1	(-1.108,-0.845)	(-0.813,2.223)	(0.203,0.337)	(-0.019, 0.019)	(0.92,1.03)
μ_2	(-0.929,-0.707)	(-0.591,1.728)	(0.178,0.342)	(-0.021, 0.017)	(0.92,1.03)
μ_3	(-1.333,-1.039)	(-1.627,-0.187)	(0.207,0.342)	(-0.031, 0.006)	(0.92,1.03)

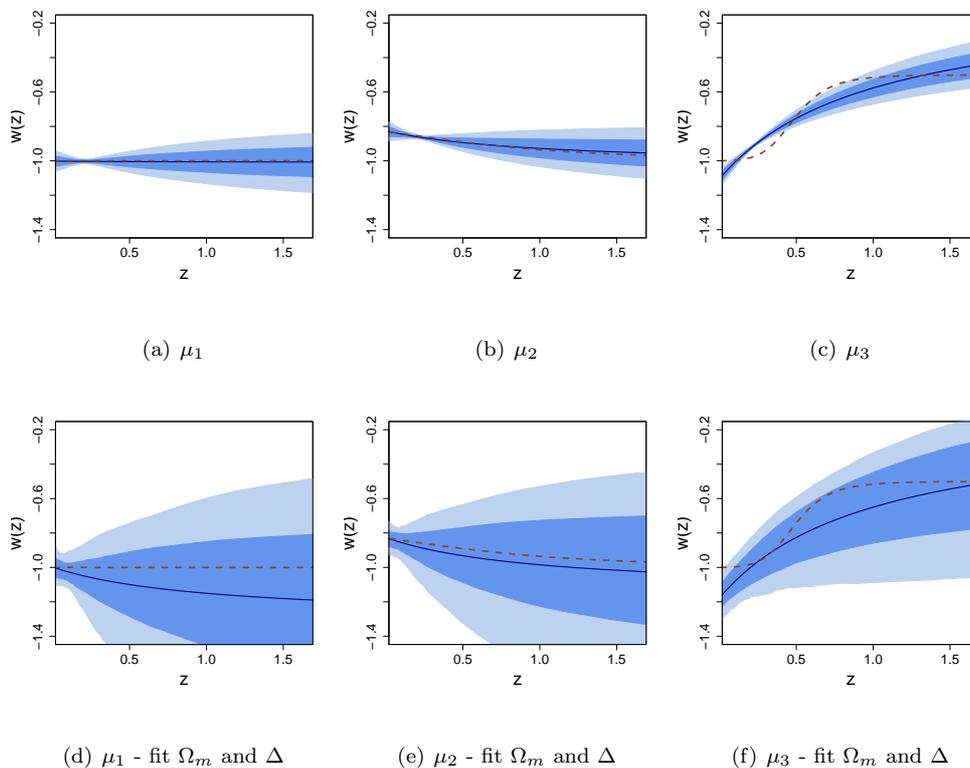


Figure 1.7: Model 3 fits of $w(z)$ for the three datasets μ_1 , μ_2 , and μ_3 . Mean (dark blue), 68% PI (blue), 95% PI (light blue), and truth (red dashed line)

Analysis

This model fits the slight curvature of simulated dataset μ_2 . This two parameter model is capable of fitting datasets μ_1 and μ_2 , however, it is not flexible enough to capture the true $w(z)$ in dataset μ_3 , as displayed in Figure 1.7. We can see the limitations of the parametric assumption for the dark energy EOS inherent in this model. But this model does an adequate job of estimating the unknown Ω_m and Δ parameters in almost all cases (Table 1.3). When these two parameters are unknown, the results have much wider probability intervals as expected. Unfortunately, there is no way to extend this model in a meaningful way to be more flexible. This seems to be the best choice of the parametric models currently in the literature.

1.3.4 Discussion on Parameter Interdependencies

The parameters in the non-linear cosmology equation are correlated to one another: specifically parameters of $w(z)$ and Ω_m . If we examine the $w(z) = a$ model's results for simulated dataset μ_3 in Figure 1.5(f), the fit of the model does not go through the truth (dashed line) and the estimate of Ω_m is not within the 95% PI bands. This fit for $w(z)$ is counter intuitive; we expect a fit to be at the mean of the truth or at least touch the true curve. The degeneracies seen in the estimates of Ω_m and Δ is a known issue. We found that there is direct relation between the degeneracy seen in some of the fits of dataset μ_3 because there are multiple $w(z)$ solutions when Ω_m is also estimated.

Basically, Ω_m and $w(z)$ produce a set of possible solutions when using SNe data alone. In Figure 1.8(a), we plot a family of curves that are equivalent solution sets to $\Omega_m = 0.27$ and $w(z) = -1$. The thick black curve is the truth. The other lines show curves of $w(z)$ for 0.1 step changes in Ω_m , which ranges from 0.16 (top curve) to 0.28 (bottom curve). We see that any curves below the true value of $\Omega_m = 0.27$ (larger Ω_m values) leads

to a degenerate $w(z)$ curve that eventually asymptotically go to negative infinity. Figure 1.8(b) has the same values of Ω_m but $w(z)$ is set to be equivalent to the truth used in simulated dataset μ_2 .

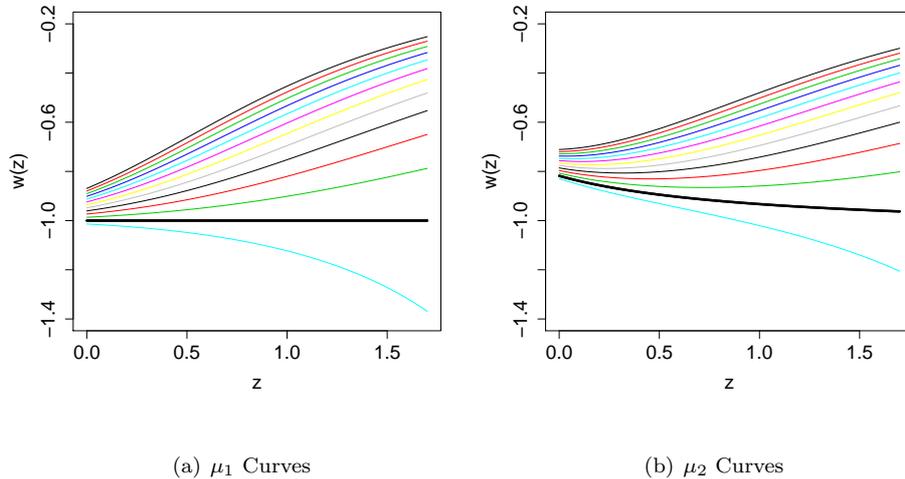


Figure 1.8: $w(z)$ curves for different Ω_m values

Figure 1.5(f) shows results for Model 1 for the highly curved simulated dataset, μ_3 . The fit of the model does not go through the truth (dashed line) and the estimate of Ω_m is also incorrect. There are multiple solutions that provide the same fit to the data with small changes in Ω_m . The plot shows that a flatter fit of $w(z)$ corresponds to a higher value of Ω_m . The SNe data alone cannot distinguish between these multiple solutions and chooses a flatter $w(z)$ curve (because of the parametric assumption being made of flatness) with higher Ω_m value. We conclude that better prior information on Ω_m is essential to retrieve appropriate and coherent fits for $w(z)$. For this reason, fits of $w(z)$ are provided for both variable and fixed Ω_m in all of the models. We try to draw strength from other observations like CMB and BAO through an informative priors for Ω_m but that does not seem to be adequate. In Chapter 3, we include CMB and BAO data directly to the analysis. These probes have different redshift-distance relationships than the SNe data and can help resolve

the multiple solutions problem of Ω_m and $w(z)$.

1.3.5 Remarks

Is the dark energy EOS equal to negative one ($w(z) = -1$)? Thus far we have presented three of the most commonly used parametric models. Currently, we have several issues in determining whether $w(z) = -1$. The underlying assumption for these three models is that the EOS for dark energy $w(z)$ follows one of the parametric forms. The one parameter model (Model 1) does not allow for a flexible form of $w(z)$ and the PI bands are quite wide on the two parameter models (Models 2 and 3). Typically, we could amend these models to relax the assumptions by adding additional terms. But adding additional terms is not possible because the combination of the data quality and complexity of the non-linear transform degrade the mean fit and PIs when anymore than two parameters are used to model $w(z)$ (Linder, 2007). The number of parameters that can be estimated is limited, which poses a real problem when trying to ascertain the functional form of the dark energy EOS.

This question could be formalized into a hypothesis test for the parametric models. For Model 1 the null hypothesis is $a = -1$ and for Model 2 and 3 would be $a = -1$ and $b = 0$. Most literature considers this a Bayesian problem with an informative prior for Ω_m and employs an MCMC algorithm to get parameter estimates. In the Bayesian framework, hypothesis testing requires point mass priors at the null hypothesis values. We did this analysis but because of the complexity of this problem the mixing on the parameter space was poor and therefore the results are not displayed. The approach we find in the literature is parametric model comparison using some type of Chi-squared test like AIC or BIC (Genovese et al., 2009). In this method, a null hypothesis model (Model 0) with $w(z) = -1$ can be compared to these other parametric models (we display the results in Chapter 2). Many one and two parameter models can quickly be compared through this method to the null

hypothesis. However, we find the standard AIC and BIC methods are unreliable for this problem, as they tend to choose the model with the least number of parameters over the correct model. The model selection criterion needs to be carefully chosen when doing model verification on a second derivative process of the data where error measures are not available. Thus we find the current methods of model fitting and model selection to be insufficient to adequately address the question, is $w(z) = -1$.

Additionally, none of the models adequately capture the shape of dataset μ_3 . We would like to have a method that allows relaxes the assumptions of the form inherent in the two parameter models and allows for a more flexible fit of dark energy EOS. We would like this method to shrink the uncertainty in the fit of $w(z)$ and have tighter PI bands. We are going to propose several non-parametric models and then propose model comparison methods that can compare parametric and non-parametric models.

Chapter 2

Non-parametric Models for $w(z)$

Back to the original question, what is the form of the dark energy EOS, could $w(z) = -1$? In Chapter 1, we present the favored parametric models from the current literature. In Chapter 2, we propose a non-parametric model that relaxes the assumptions about the form of the dark energy EOS. The parametric methods start with a guess at the form of $w(z)$, typically via two parameter models, then implement and analyze them all, and conclude with a model comparison method to choose the best one. The form of $w(z)$ is limited to at most a two parameter model and the current model comparison techniques may not be providing accurate information as to which two parameter form is best. Unfortunately, adding any higher order terms to the model for $w(z)$ degrades the fit and the uncertainty increases to a point where the fit is meaningless. Most non-parametric methods rely on higher order terms or an expansion of basis functions (Genovese et al., 2009). A non-parametric fit is of great benefit because it is flexible and can model any continuous form of the dark energy EOS.

In this chapter, we employ two non-parametric models to do inference on $w(z)$, we call these Models 4 and 5. We continue to use the $r(z)$ equation as our distance-redshift relationship and the same simulated SNe dataset. In Model 4, we assume $w(z)$ is a Gaussian

process (GP). A GP is a stochastic process such that when sampled at any finite collection of points, the values jointly follow a multivariate Normal distribution (Banerjee et al., 2004; Rasmussen and Williams, 2006). Thus the process can be fully defined by its mean and correlation functions. This model is advantageous in that it allows for a flexible fit to a function, $w(z)$, based on probability theory rather than assuming a parametric form. We begin with an general example of modeling a derivative process of interest with a GP that is not related to cosmology. This example is to explain the inverse method we develop and compare it to other GP methods typically used in the statistical literature. Then we apply our GP method to the cosmology problem in Model 4. Next we show Model 5, which is an approximation to a GP comprised of a basis of damped Hermite polynomials (Steinberg and Burstzryn, 2004). These models give a non-parametric fit and an approximation to a non-parametric fit for $w(z)$.

Other non-parametric modeling options for $w(z)$ exist but we do not explore them here. These include a piecewise constant model that requires binning of the SNe data into categories based on their redshift values, called principal component analysis (PCA) (Huterer and Starkman, 2003). It has some benefits in that it relies on the data to weight parts of the model and can focus on analyzing specific redshift values. But it produces a discrete step function estimate for $w(z)$ which does not adhere to the physics of the problem (Krauss et al., 2007; Crittenden and Pogosian, 2005; Simpson and Bridle, 2006). Other approaches include using parametric forms of $w(z)$ based on higher order polynomials, which tend to break down (Genovese et al., 2009; Linder, 2007). The non-parametric methods we present assume a continuous $w(z)$ unlike many of the current non-parametric approaches like PCA. And the GP method is flexible in its form without degrading the fit like higher order polynomial models.

2.1 Gaussian Process Models

For the non-parametric GP, $w(z)$ is modeled with a collection $w(z_1), \dots, w(z_n)$, for any set of z_1, \dots, z_n , which follow a multivariate Gaussian distribution with mean, θ , and powered exponential covariance function (Neal, 1997). Thus $w(z)$ is considered a GP: $w(z) \sim GP(\theta, \kappa^2 K(z, z'))$. A standard powered exponential covariance function suffices but we re-parameterize the typical form to $\Sigma(z, z') = \rho^{|z-z'|^\alpha}$ where $\rho = e^{-\lambda}$ and $\rho \in [0, 1)$. This form is used because ρ is a smoothness parameter in the model and a prior can be better formed in this way. For the exponential correlation function, a ρ value near one means the GP is rather smooth and a ρ value near zero would have a much rougher GP realization. The underlying question of whether $w(z)$ is constant and equal to negative one means we want to know if ρ is approaching its limit as $\rho \leftarrow 1$.

2.1.1 Integration of a GP

We are interested in $w(z)$, a derivative process. Typically, this type of problem is approached in the opposite manner where one fits the data directly with a GP and then takes the derivative of the GP which results in a new GP (Banerjee et al., 2004). However, we are most interested in the derivative process and not modeling the data; it has been shown that it is best to model the equation of interest directly. Therefore, we view this as an inverse problem, as we previously did with the parametric models. Inverse problems can be complicated with identifiability issues because the likelihoods may not be written in closed form (Higdon et al., 2003; Kaipio and Somersalo, 2004). We use the logic of the current derivative GP method but instead invert it, showing that the integral of a GP is itself a GP. This is a method similar to Bayes-Hermite quadrature described by O'Hagan (1991) where the GP, $w(z)$, is integrated based on the properties of the GP. However, O'Hagan (1991) applies this method to integrating a continuous function; we adapt this method to

integrate a stochastic process. We have a derivative process, $w(u)$, we wish to model directly by assuming it is a non-parametric GP with m sampled points (as seen in equation (2.1)). We need to evaluate $y(s) = \int_0^s w(u)du$ and show that based on the properties of the GP, $y(s)$ is also a GP (as seen in equation (2.2)). Equation (2.3) shows the relationship between the two GPs: $w(u)$ and $y(s)$ (for a more extensive proof see Appendix A.1).

$$w(u) \sim GP(\theta, \Sigma_{22} = \kappa^2 \rho^{|u-u'|^\alpha}) \quad (2.1)$$

$$y(s) \sim GP\left(\theta s, \Sigma_{11} = \kappa^2 \int_0^s \int_0^{s'} \rho^{|u-u'|^\alpha} du' du\right) \quad (2.2)$$

$$\begin{bmatrix} y(s) \\ w(u) \end{bmatrix} \sim MVN \left[\begin{bmatrix} \theta s \\ \theta \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right] \quad \text{where } \Sigma_{12} = \kappa^2 \int_0^{s'} \rho^{|u'-s|^\alpha} du' \quad (2.3)$$

This results in a distribution for $y(s)|w(u)$. We only need a single form for each $y(s)$ and not a whole distribution. O'Hagan (1991) shows that $y(s) \approx E(y(s)|w(u))$ is sufficient as an answer to the numerical integration when compared with other numerical integration methods and the $V(y(s)|w(u))$ term is not taken into account. After some analysis, we found that $V(y(s)|w(u))$ is sufficiently small (no more than noise) when we have more than eight sampled points in $w(u)$. We ran the algorithm with fifty to one hundred GP points for $w(u)$, which is more than a sufficient amount for $V(y(s)|w(u))$ to be close to zero.

In the same vein, Yaglom (1962) suggests that GPs be integrated by a basic approximation method, rectangle integration. A grid (ν) over the domain (z) has sampled GP points ($w(z)$). For an exact solution, this is an infinite grid but in practice a finite grid is used, where enough grid points supplies a good estimate of the integral /citepBan08. It can be shown that a finite sum of the GP points that comprise $w(u)$ gives a one-to-one relationship with the GP points of $y(s)$ which is the expectation seen in equations (2.4)-(2.6). We

perform numerical integration for $y(s)$ using equation (2.7).

$$y(s)|w(u) = \sum_{i=1}^{\infty} w(\nu_i)(\nu_i - \nu_{i-1}) \quad (2.4)$$

$$\approx \sum_{i=1}^m w(\nu_i)(\nu_i - \nu_{i-1}) \quad (2.5)$$

$$= E(y(s)|w(u)) \quad (2.6)$$

$$= \theta s + \Sigma_{12}\Sigma_{22}^{-1}(w(u) - \theta) \quad (2.7)$$

The only computationally intensive calculation needed is Σ_{12} which requires a single integral for every entry. This integration cannot always be done analytically because it depends on the correlation function; we use the Chebyshev-Gauss quadrature method for numerical integration, see Appendix A.2.

2.1.2 General Example

We begin with a generic example to compare this method to other GP methods for finding derivative processes. We generate $n = 100$ equally spaced data points from the curve $f(z) = \ln(1+z)$ for $z \in [0, 5]$ and add noise $\epsilon \sim N(0, 0.5^2)$. Thus, the derivative curve is $f'(z) = 1/(1+z)$. We start by fitting a GP directly to the data and then differentiate it to obtain an estimate of the derivative curve. This is the typical approach used for GP modeling of a first derivative process (Neal, 1997). Banerjee et al. (2004) suggests using a Matérn with smoothness parameter $\nu=3/2$ (or $\nu \in (1, 2)$ if you want to fit the ν parameter), parameterized as $cov(w(z), w(z')) = \kappa^2(1 - \log(\rho)|z - z'|)\rho^{|z-z'|}$, where $\kappa > 0$ is a scaling parameter and $\rho \in (0, 1)$ is a correlation parameter. This is one of the very few correlation functions that has the appropriate properties to make differentiation of a GP possible. It is once differentiable but it does not have the numerical issues like the overly smooth Gaussian or the other Matérn correlation with a higher smoothness parameter. These correlation functions would require a nugget (jitter) term because of numerical issues but

this is theoretically incorrect to use when differentiating. We show that the differentiation approach is an inferior method when compared to modeling the derivative process directly and viewing this as an inverse problem. In this example, we focus on the power exponential family. We choose this form because we want to compare the differentiation approach versus the inverse approach. Then we move forward and apply the best method to the cosmology problem. The Matérn ($\nu = 3/2$) is too smooth for the inverse method in the cosmology application. We use the power exponential family, $\alpha \in (0, 2]$ is typically equal to two (Gaussian correlation) but this leads to numerical singularities in the matrix. One option is to add a nugget term or jitter but the prediction of the derivative process is degraded (Stein, 1999). Instead, we incorporate the observational error into the likelihood equation in our model, much like we did with the parametric models. Cholesky decomposition with pivoting was considered to deal with the numerical instability of the covariance matrix. Ultimately, we let α be slightly less than two; in most cases the approximation 1.9999 works well (we refer to this as the pseudo-Gaussian correlation function).

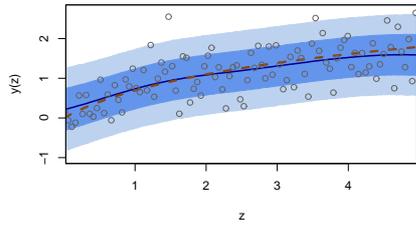
In order to have differentiability for the first method, we can only use a Gaussian correlation function from the exponential family. Our second fit corresponds to the derivative approach based on $y(z) \sim GP\left(0, \kappa^2 \int_0^z \int_0^{z'} k(u, u'; (\rho, \kappa)) du du'\right)$. In this case, we use $\alpha = 1$, so that the double integral for the correlation function for $y(z)$ can be calculated in closed form. The third fit is obtained using the inverse method we present in the previous section using $E(w(u)|y(s))$ as our transform. We show our approach with both the exponential correlation function ($\alpha = 1$) in the third method and the Gaussian correlation function ($\alpha = 1.9999$) in the fourth method. For the Gaussian correlation, we compute the integrated correlation needed for the matrix K_{yw} numerically, using Gauss-Chebyshev quadrature detailed in Appendix A.2.

In all methods, we use the same priors while generally avoiding non-informative priors because of potential issues with improper posteriors. The parameter ρ is defined on

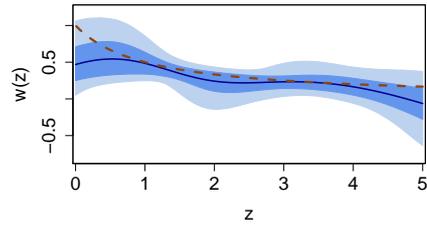
the interval $(0, 1)$. We desire smooth GP realizations that have higher correlation values, therefore, we have an informative prior on ρ that favors values close to one: $\pi(\rho) \sim Be(6, 1)$. σ^2 is the variance of the observational error, a conjugate non-informative prior just like in the parametric models: $\pi(\sigma^2) \sim \sigma^{-2}$. κ^2 controls the variance of the GP; we want an informative prior with values away from zero, as this would produce numerical instability: $\pi(\kappa^2) \sim IG(4, 1)$. The last parameter to consider is θ ; we assume it to be constant which is equivalent to the conventional method of subtracting off the data mean of the derivative process or a linear trend on the data scale. For this particular problem we have set $\theta = 0.41$. We use an MCMC algorithm based on a combination of Gibbs and Metropolis steps. To ensure good convergence and mixing for the posterior chains we run the chains for 100,000 iterations.

The results for all four fits are presented in Figure 2.1. Panel 2.1(a) shows the curve fitted by using a GP directly and 2.1(b) shows the resulting estimation of the curve derivative. Even though the point-wise PIs cover the true values of the derivative for most of the range, it is clear that the estimation is very wavy. Furthermore, the quality of the fit is poor for values of z in the extremes of the interval. The results obtained for the estimation of the derivative curve using the inverse problem approach are superior to the direct method in all cases.

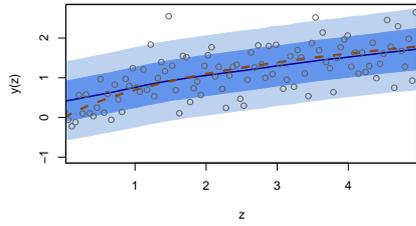
A comparison of the panels in the left column reveals that the method based on the approximate likelihood produces intervals for the data fitting curve that are slightly narrower than the exact method. This is not surprising, as the approximation ignores some of the variability in the integration of the latent process $w(z)$. On the other hand, the panels on the right column indicate that the derivative curve is estimated quite accurately. This is particularly in the case of the fit in panel 2.1(h), corresponding to the approximated likelihood method with Gaussian correlation function. Since, the function of interest is not the data curve but rather the derivative process, we find that our inverse method and



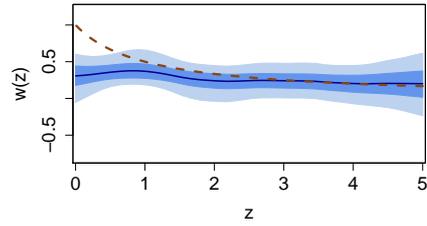
(a) Direct data fit



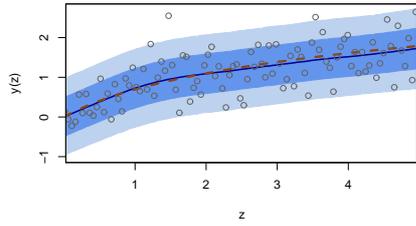
(b) Induced estimation of f'



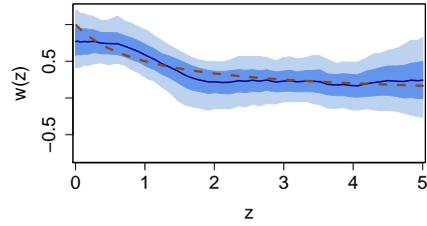
(c) Direct data fit, $\alpha = 1$



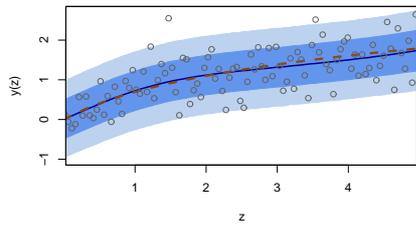
(d) Induced estimate of f' , $\alpha = 1$



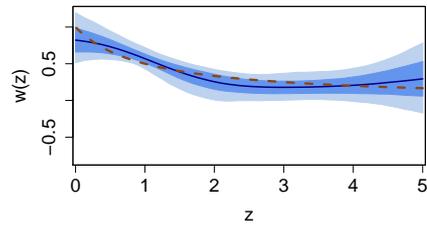
(e) Approx. inv. data fit, $\alpha = 1$



(f) Approx. inv. f' estimation, $\alpha = 1$



(g) Approx. inv. data fit, $\alpha = 1.99$



(h) Approx. inv. f' estimation, $\alpha = 1.99$

Figure 2.1: Four estimation approaches for $f'(z)$. Left column: simulated data (circles); true curve (dashed line); fitted curve (solid line). Right column: true derivative curve (solid line); estimated derivative (dashed line). 65% and 95% probability interval bands (dark blue and light blue respectively).

integration technique for the GP seems to fit the best. We wish to apply this GP method to the cosmology example.

2.2 Non-Parametric Models for the Dark Energy EOS

2.2.1 Model 4 - $w(z) \sim \text{GP}$

We intend to apply the inverse GP method to the cosmology problem of fitting the functional form of the unknown dark energy EOS, $w(z)$. We use a standard powered exponential covariance function: $\Sigma(z, z') = \rho^{|z-z'|^\alpha}$. The pseudo-Gaussian correlation and the Matérn correlation (with smoothness parameter $\nu=1.5$) both assume a very smooth functional form for the dark energy EOS ($w(z)$). We find that the exponential correlation ($\alpha = 1$) leads to a much more flexible model that fit both flat and curved $w(z)$ equations and can capture the extreme curvature seen in dataset μ_3 . To ensure over-smoothing is not happening a two step process is adopted. First, $\alpha = 1$ model is fit and if the resulting $w(z)$ is relatively flat, we fit a model with $\alpha = 1.9999$ without concern of over-smoothing.

We assume $w(z)$ is a GP (as seen in equation (2.8)) with m sampled points. We must evaluate $y(s) = \int_0^s \frac{w(u)}{1+u} du$ based on the properties of the GP and obtain $y(s)$, a non-stationary GP (as seen in equation (2.9)). Equation (2.10) is the relationship between the two GPs: $w(u)$ and $y(s)$ (see Appendix A.1 for more details).

$$w(u) \sim GP(\theta, \Sigma_{22} = \kappa^2 \rho^{|u-u'|^\alpha}) \quad (2.8)$$

$$y(s) \sim GP\left(\theta \ln(1+s), \Sigma_{11} = \kappa^2 \int_0^s \int_0^{s'} \frac{\rho^{|u-u'|^\alpha}}{(1+u)(1+u')} du' du\right) \quad (2.9)$$

$$\begin{bmatrix} y(s) \\ w(u) \end{bmatrix} \sim MVN \left[\begin{bmatrix} \theta \ln(1+s) \\ \theta \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right] \quad (2.10)$$

We run an algorithm with fifty GP points for $w(u)$, which should be more than a sufficient amount for a good numerical approximation. $y(s)|w(u)$ is a distribution and we

perform numerical integration for it as follows:

$$y(s)|w(u) = E(y(s)|w(u)) = \theta \ln(1 + s) + \Sigma_{12} \Sigma_{22}^{-1} (w(u) - \theta)$$

This handles just the inner integration step, now we need an effective outer integration procedure. The outer integration is done by a numerical trapezoid method (because of the non-linear transform) and requires many grid points for an accurate numerical integration. This is actually n integrations, one for each data point. We draw more $y(s)$ points than are in original partition of $w(u)$ by adding interpolated points, which gives enough points for a more precise outer integral calculation. This method does not require a large covariance matrix to be computed and inverted because the extra smoothing points are being sampled only for the outer integration. Additionally, the inner integral and partitioning/smoothing process are all done in one step, which is quite efficient. Σ_{11} is never computed so the slowness of computing double integrals for each entry of the covariance matrix is avoided. However, we need to compute $\Sigma_{12} = \kappa^2 \int_0^{s'} \frac{\rho^{|u-s|^\alpha}}{1+u} du$ which requires a single integral for every entry, see Appendix A.2 for details on numerical integration.

There are other numerical issues when using a GP model embedded in a highly non-linear transformation. Usually, when using GPs the posterior distributions result in closed form representations that can be drawn with Gibbs steps. This inverse GP model with m points must be sampled with a Metropolis algorithm at every iteration because there is no closed form posterior distribution (but this was also true for the parametric models). This step of the MCMC algorithm tends to have a high rejection level because we propose m new sampled points in a joint step. To reduce the rejection level, the likelihood is rewritten into an equivalent form that allows for smaller steps between iterations. This is a more localized search algorithm used for GPs and is presented in Appendix A.3. The full computational MCMC algorithm for the GP method is laid out in detail in Appendix A.4 and see Appendix A.5 for a further discussion on MCMC issues.

There are four parameters that comprise the mean and covariance functions of the GP: α , ρ , τ , and θ . α is either set to 1 or 1.9999, exponential or Gaussian correlation respectively. The mean function of the GP has parameter θ . Allowing θ to be an unknown variable did not produce stable posterior results in this analysis. We start by setting θ equal to negative one and updated it in an iterative manner based on the posterior mean of the GP after a few runs. This is very similar to subtracting off a data mean and assuming the trend of interest to be a zero mean GP. The other parameters are ρ and κ in the correlation function. ρ is considered the correlation length but does not carry a straight-forward physical interpretation. We found that α actually controls most of the smoothness properties of the GP, while ρ has much less influence than expected. κ controls some of the variability around the mean (θ) but also is highly correlated to ρ . We are not concerned with the physical meaning of ρ and κ as they are interdependent and both are estimated, not fixed.

All of the unknown parameters in the model need priors. We give ρ a more informative prior towards one ($\pi(\rho) \sim \text{Beta}(6, 1)$) which assumes more smoothness in the fit of $w(z)$ because we are using the rougher exponential correlation function. A non-informative prior on τ results in an unstable posterior, so τ is given an informative prior with as low weight as possible: $\pi(\kappa^2) \sim \text{IG}(6, 2)$. σ^2 has a non-informative prior: $\pi(\sigma^2) \propto \sigma^{-2}$, where σ^2 is the observational variance discussed in Chapter 1. σ^2 is treated the same in both the parametric and non-parametric models.

The two additional parameters, Ω_m and Δ are initially fixed, then in the second set of simulations are considered as variable parameters just as in the parametric models. The results of the GP model for $w(z)$ are displayed in Figure 2.1. All sets were run 200,000 iterations. The mean value for the GP prior is set based on the posterior mean of an initializing run of the GP. The mean is set at -1 for all dataset μ_1 runs. Dataset μ_2 has its mean set at -0.94, and dataset μ_3 has its mean set at -0.7. When Ω_m and Δ are variable the mean for dataset μ_2 is set at -0.87 and for dataset μ_3 , it was set at -1.

Results

The best way to summarize the results of the GP model of $w(z)$ is in figures, which can also be compared with the plots of the fits of the parametric models shown in Chapter 1. Figure 2.2 shows the GP model is capable of finding the true $w(z)$ for all three simulated datasets: μ_1 , μ_2 , and μ_3 . The roughness in the plots comes from the exponential correlation function. All other correlation functions over-smooth the fit and do not allow enough flexibility for fitting a true $w(z)$ in simulated dataset μ_3 . If we find that the fit for $w(z)$ is not highly curved in the GP model with exponential correlation then we can fit a GP with pseudo-Gaussian correlation without concern that this model will over-smooth. Simulated datasets μ_1 and μ_2 can be fit with a GP with pseudo-Gaussian correlation function but not dataset μ_3 .

Table 2.1: Model 4 - Posterior 95% PIs

Dataset	Ω_m	Δ
μ_1	(0.227,0.302)	(-0.021, 0.016)
μ_2	(0.212,0.309)	(-0.022, 0.014)
μ_3	(0.264,0.365)	(-0.026, 0.014)

We have presented a non-parametric model that captures the curvature of $w(z)$, while not introducing larger probability intervals than the two-parameter models. Overall, the GP fits well when Ω_m is known. However, when Ω_m is allowed to vary, we see that the true Ω_m is at the boundary of the estimated interval in datasets μ_1 and μ_3 . This is also seen in several of the parametric models discussed in Chapter 1 and explored in section 1.3.4 about parameter interdependencies. This is a known issue with multiple solutions of the cosmological model; to abate this issue more information (possibly in the form of a tighter prior from other data sources) is needed for Ω_m . Even though, we have a novel non-parametric approach it alone cannot answer if $w(z) = -1$ until more information is available to better constrain Ω_m ((Holsclaw et al., 2010b)). Chapter 3 is an investigation

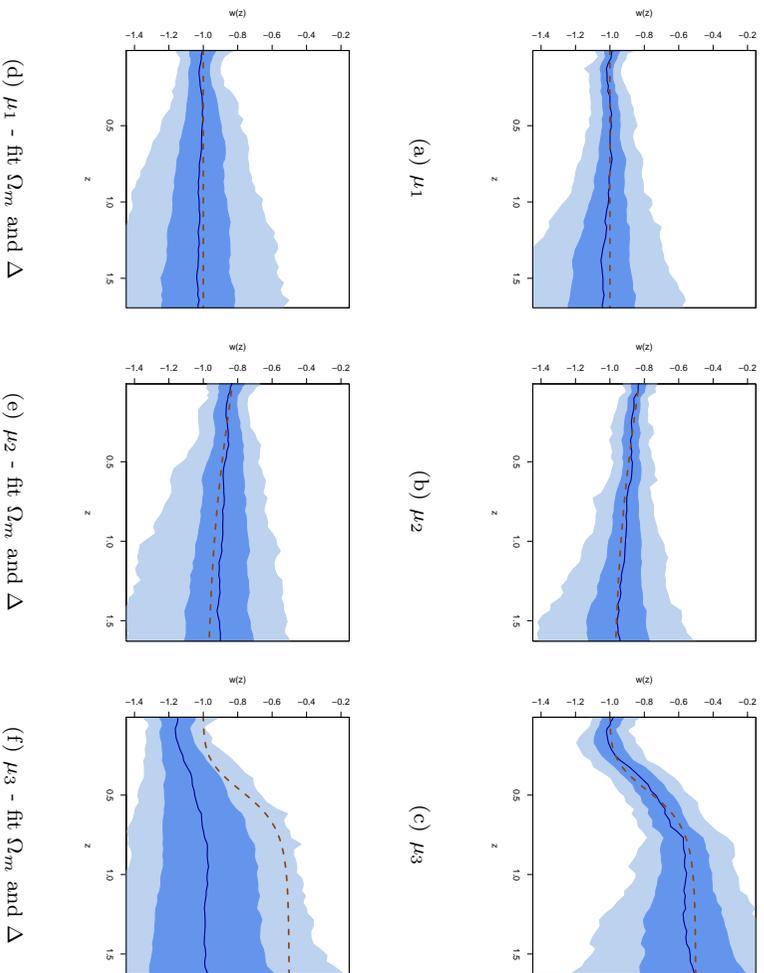


Figure 2.2: Model 4 fits of $w(z)$ for the three datasets μ_1 , μ_2 , and μ_3 . Mean (dark blue), 68% PI (blue), 95% PI (light blue), and truth (red dashed line)

of additional data sources that can shrink the uncertainty associated with Ω_m which would lead to a more accurate prediction of the form of $w(z)$. In Chapter 3, instead of having a prior for Ω_m based on other data sources, we introduce the data sources directly through their redshift-distance relations to abate the issues of multiple solutions for Ω_m and $w(z)$.

2.2.2 Model 5 - $w(z)$ is Modeled by a Damped Hermite Polynomial Basis

Before moving on to multiple data sources, we investigate a model that approximates a GP with a basis of damped Hermite polynomials, this follows the work of Steinberg and Burstzyn (2004). A damped polynomial basis applied to $w(z)$ approximates a GP model. Linder (2007) opposes fitting $w(z)$ with higher order polynomials (beyond two terms) because of their oscillating behavior and the need to truncate the fit for higher z values, as well as, the fact that they can introduce bias in the fit. However, Genovese et al. (2009) use a mixture of basis functions (which need not be orthonormal) to fit $w(z)$. Their analysis allows for three forms that include polynomial basis, scale factor polynomials, and piecewise constant fits and can be extended to include B-splines, orthogonal polynomials and wavelet analysis (Genovese et al., 2009). However, we believe after reproducing these results that they may have only tried the first few orders of the basis function in their comparison and had the same issues as Linder (2007) when using anything beyond two parameter models for $w(z)$. Using model selection criteria (BIC), they find that when using the polynomial expansions the best model for $w(z)$ is a constant model. We show that these results may not be valid in section 2.3 as the BIC is overly conservative.

We found models with more than two parameters degrade; even when using the damped Hermite polynomial expansion. We are only able to fit up to two polynomials in a stable manner. We examine the damped polynomials here because they are a different form of polynomials than has been previously used for this problem. We let $W(u) = \frac{w(u)}{1+u}$ and

we fit a series expansion of basis functions to $W(u)$, shown in equation (2.11).

$$W(u) = \gamma(u) + \sum_{s=0}^{\infty} \beta_s J_s(u) \quad (2.11)$$

where the mean of $W(u)$ is $\gamma(u) = \frac{-\theta}{1+u}$, θ is a constant, $J_s(u) = H_s^*(u)e^{-\frac{mu^2}{2(1+m)}}$ are the damped polynomials, and $H_s^*(u) = H_s(u/\sqrt{(2)})/(2^s s!)^{1/2}$ is a physicist Hermite polynomial. The Hermite polynomials, H_s^* , are orthonormal with respect to a standard Normal, and have the property:

$$E[H_s^*(W)H_t^*(W)] = \int_{-\infty}^{\infty} \exp^{-u^2/2} H_s^*(u)H_t^*(u)du = \delta_{s,t}$$

But the damped polynomials $J_s(u)$ that comprise the approximating summation are not orthogonal.

Integrating the series expansion that approximates $W(u)$ is not tractable. We choose to use $W(u)$ instead of $w(u)$ because this allows the inner integration in the transform to be performed using the properties of the Normal distribution, which is well approximated by our statistical software and is quite computationally efficient. The integration method is shown in equations (2.12) - (2.14).

$$\int_0^s W(u)du = \int_0^s \gamma(u)du + e^{-\frac{mu^2}{2(1+m)}} (a_0 + a_1u + a_2u^2) du \quad (2.12)$$

$$= \int_0^s \frac{-\theta}{1+u} du + a_0 \int_0^s e^{-\frac{u^2}{2k^2}} ds + a_1 \int_0^s ue^{-\frac{u^2}{2k^2}} du + a_2 \int_0^s u^2 e^{-\frac{u^2}{2k^2}} du \quad (2.13)$$

$$= -\theta \ln(1+s) + a_0 \sqrt{2\pi k^2} [N(s|0, k^2) - 1/2] + a_1 \{-k^2 e^{-s^2/2k^2} + k^2\} \quad (2.14)$$

$$+ a_2 \{-k^2 e^{-s^2/2k^2} (s) + k^2 \sqrt{2\pi k^2} [N(s|0, k^2) - 1/2]\}$$

The damped Hermite polynomial parameterization approximates a GP with Gaussian correlation with range $\lambda = \frac{m}{2(1-m^2)}$ and variance $\sigma^2 = \tau^2(1-m^2)^{-1/2}$ and $0 \leq m \leq 1$. Because the redshift axis (z) ranges between zero and two, the basis of Hermite polynomials needs to be rescaled by a factor of two or three ($J(2z)$ or $J(3z)$). The unscaled polynomials do not dampen until closer to $z = 3$ which is beyond the range of

data. Both τ^2 and m are unknown parameters that must be estimated, which have priors: $\pi(m) \sim \text{Beta}(6, 1)$ and $\pi(\tau^2) \sim \text{IG}(25, 9)$. The prior on β also contains m , the damping parameter, $\pi(\beta_s) \sim \text{MVN}(0, \sigma^2 m^s)$. The observational variance term, σ^2 , in the likelihood has a prior of: $\pi(\sigma^2) \sim \sigma^{-2}$.

We write the series expansion up to the first two polynomials with β 's and show that when the polynomials are truncated they can be expressed with new coefficients, a 's.

$$W(u) = \gamma(u) + \sum_{s=0}^{\infty} \beta_s J_s(u) \quad (2.15)$$

$$= \gamma(u) + \exp\left(-\frac{mu^2}{2(1+m)}\right) \left[\beta_0 - \frac{1}{\sqrt{2}}\beta_2 + \beta_1 u + \left(\beta_2 \frac{1}{\sqrt{2}}\right)u^2 \dots \right] \quad (2.16)$$

$$= \gamma(u) + \exp\left(-\frac{mu^2}{2(1+m)}\right) \left[a_0 + a_1 u + a_2 u^2 \dots \right] \quad (2.17)$$

In Figure 2.3, we see the first six damped Hermite polynomials. Figure 2.3(a) is of the H^* polynomials corresponding to the β_s coefficients and Figure 2.3(b) shows the polynomial basis corresponding to the a_s coefficients. We set $m = 0.9$ for these graphs and use a rescaling factor of three.

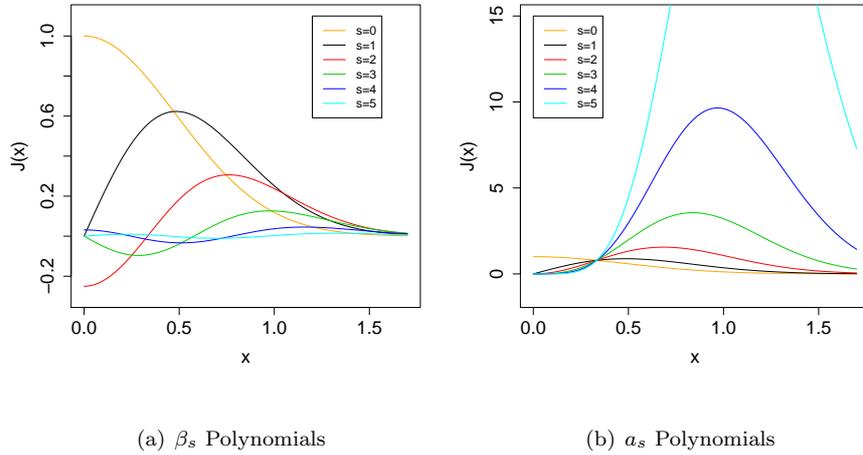


Figure 2.3: Damped Hermite Polynomials

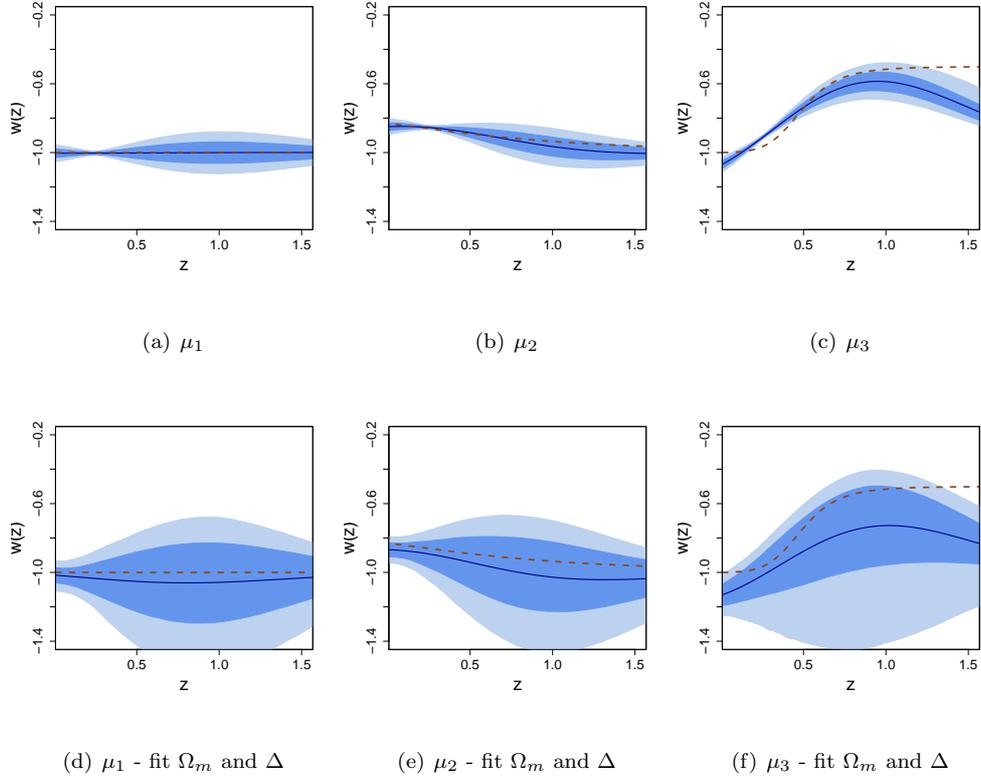


Figure 2.4: Model 5 fits of $w(z)$ for the three datasets μ_1 , μ_2 , and μ_3 . Mean (dark blue), 68% PI (blue), 95% PI (light blue), and truth (red dashed line)

Table 2.2: Model 5 - Posterior 95% PIs

Dataset	Ω_m	Δ
μ_1	(0.218,0.322)	(-0.022,0.014)
μ_2	(0.216,0.332)	(-0.022,0.012)
μ_3	(0.225,0.368)	(-0.026,0.010)

Results

As with all basis expansions for this problem, only the first two terms of the damped Hermite polynomial expansion are used before the analysis degrades. Thus we prefer the actual GP model over this approximation method for the GP. We hoped this method would be more computationally efficient than the actual GP model. But the GP model has been creatively sped up in such a way that its performance is comparable from an efficiency standpoint. This model approximates a GP with Gaussian correlation which does not fit dataset μ_3 well because it over smooths as expected. But Table 2.2 shows the true values of Ω_m and Δ are in the PIs.

2.2.3 Conclusions

Parametric models typically result in more constrained parameter estimates than those from non-parametric models. But parametric models require a form to be chosen a priori; typically many forms can be fit and then compared through a model comparison technique without knowing if any of the forms are correct. In this application, the parameterization of $w(z)$ is limited to forms no higher than a first order polynomial fit, which is very constraining. Genovese et al. (2009) work with a basis of functions, which they claim cover the polynomial and piecewise constant cases. We replicated the polynomial basis expansion (not shown in this document) and found it degraded quickly as more terms were added. We have pursued some different non-parametric models in the hope of giving a better alternative to the inflexible parametric approaches or non-parametric methods that rely on binning. The inverse GP method we present provides narrower probability bands than some of the frequently used two-parameter models.

When assuming $w(z)$ is a GP, this gave preferable results to other models for determining the form of the dark energy EOS. Not only is it continuous and flexible (because

of our choice of correlation function) but it can model a large range of possible $w(z)$ forms and it has smaller PI bands than many of the two parameter models. But the GP model is computationally more expensive. By using the distribution properties of the GP, we can speed up the computations considerably. However, the uncertainty associated with the parameter Ω_m is problematic when determining the truth about $w(z)$ (we want to work to resolve this in Chapter 3). We also desire a method of model comparison that is more reliable than what has been presented in the literature for this problem. Overall, we want a way to quantify which model is best and if we can determine if $w(z) = -1$.

2.3 Hypothesis Testing and Model Comparison

2.3.1 Hypothesis Testing

When comparing parametric models there are standard methods of comparison and hypothesis testing for a null hypothesis of $w(z) = -1$ (Genovese et al., 2009). But how do we compare parametric models versus non-parametric models? We need something special to do hypothesis testing and model comparison because some of the models are parametric and others are non-parametric.

Here we view hypothesis testing as a part of model comparison. We set the null hypothesis to be $w = -1$ and run a model (Model 0). Then we use model comparison methods to compare this null hypothesis model to all the other models. Other methods of Bayesian hypothesis testing using point mass priors and a reversible jump type algorithm did not work well with this problem. Typically, these types of algorithms are for posteriors with closed form; in this problem those are unavailable. The convergence of the posterior chains in the algorithm are questionable and highly dependent upon the proposal distributions in the algorithm. This is unfavorable and thus the results questionable (see Appendix A.5 for a further discussion on MCMC issues).

2.3.2 Model Comparison

For parametric model comparison (Genovese et al., 2009) uses a Chi-squared test (Bayesian Information Criterion (BIC)) to choose between competing models and concludes that the simple $w = a$ model is the best (Schwarz, 1978; Szydlowski and Wlodzimierz, 2006). We test the BIC and AIC on our simulated data and they tend to choose models with too few parameters. This test is too cautious against over parameterizing and thus is unreliable. We assume the BIC is failing because the test is meant to be used on the level of the data (μ) and not on the embedded second derivative process of a non-linear function like $w(z)$. Residual analysis has the same issues as they are on the level of the data and meant to assess a fit of the data and not the non-linearly transformed second derivative process. We have found that if the wrong parametric model is chosen for $w(z)$ the residuals may not show any deviation in the residuals of the data. Currently, these are the methods being used to compare multiple parametric possibilities for $w(z)$. There is no way to know if a particular parametric model is the correctly assumed form using these traditional methods.

We want a method of model comparison that provides coherent results for the different forms of $w(z)$. We tried several Bayesian model comparison that did not perform well. One is Bayes factors (BF) which are difficult because of the complexity of the model and require an approximation algorithm to work (March et al., 2010). Moreover they are highly sensitive to the non-informative or lowly informative priors included in all of the parametric models. Overall, BF are not known for working properly unless all priors are somewhat informative; the results of the BF tests in this application are highly influenced by the priors and therefore we avoid displaying the results because with different priors we get different results of the model comparison. Also, it is not completely clear how to treat the 50 finite GP points in the non-parametric model. We can choose to sample any number of points for the GP; we use fifty grid points and they would need each be a parameter in the

BF. Between the issues with the priors and the approximation algorithms the BF method did not work well for this application.

The posterior predicted loss (PPL) criterion and the deviance information criterion (DIC), both seemed to work better than any of the other methods (Banerjee et al., 2004; Gelman et al., 2004; Gelfand and Ghosh, 1998). They also work for non-parametric models, which is essential since we have presented a GP model. We tried both the quadratic and linear loss for the PPL method (only the results for linear loss are presented in Table 2.3 because the two loss functions were quite similar). The DIC is defined as $DIC = 2\bar{D}(\psi) - D(\bar{\psi})$, where ψ is the set of unknown parameters in the model, $D(\psi) = -2\log(p(y|\psi))$ and $\bar{D}(\psi)$ is the average of each iterations $D(\psi)$ value in the MCMC and $D(\bar{\psi})$ is $D(\psi)$ evaluated at the average value of ψ from the iterations. Overall, we found the DIC to have some added benefits like being able to estimate the number of effective parameters ($p_D = \bar{D}(\psi) - D(\bar{\psi})$) in the models.

In a parametric model, the number of parameters are counted in a straight-forward manner but in the non-parametric GP, there is no count for the number of parameters (this is the reason the AIC and BIC are not options, as they require the number of parameters). The DIC produces an estimated effective number of parameters as part of the test (Spiegelhalter et al., 2002). This value is of interest because some parameters are correlated and do not effectively give a full count of one and more informative priors also shrink the effectiveness of a parameter. We do not expect in parametric models that the full count of the number of parameters equals the effective number of parameters because of the informative priors and correlation among parameters.

We notice when the DIC is calculated for the simulated datasets, it chooses the correct model often but the values of the DIC are quite close together. Once again, this could be due to the fact, we are running the DIC on the second derivative process fit, $w(z)$, and not directly on a model fit to the data (μ). The DIC is the method that performs the best

on these simulated datasets. We perform each model comparison test on the null hypothesis model ($w(z) = -1$) and all of the parametric and non-parametric models presented thus far in Table 2.3. For a quick comparison, we note each model with correct form for the simulated data with a star and then the model every test chooses with a star. The minimum value of each test is the favored model.

Table 2.3: Model Comparison - Fixed Δ and Ω_m

Data	Model	BIC	AIC	p	DIC	p_d	PPL
μ_1	0*	-2922*	-2928*	1	-2928.0*	1.00	427.6094
	1	-2915	-2926	2	-2926.4	1.93	427.4711
	2	-2907	-2924	3	-2924.3	2.96	427.5175
	3	-2907	-2924	3	-2924.0	3.10	427.6386
	4	n/a	n/a	n/a	-2924.6	3.00	427.4007*
	5	-2892	-2920	5	-2924.1	3.06	427.4983
μ_2	0	-2483	-2489	1	-2489.2	1.00	470.5905
	1	-2913*	-2925*	2	-2924.9	1.97	427.4999
	2	-2907	-2924	3	-2924.3	2.97	427.4614
	3*	-2907	-2924	3	-2924.2	3.05	427.4943
	4*	n/a	n/a	n/a	-2924.2	2.97	427.4522*
	5*	-2891	-2920	5	-2927.5*	-0.68	427.5024
μ_3	0	-2713	-2719	1	-2719.1	1.01	448.0934
	1	-2871	-2882	2	-2882.4	2.05	431.6029
	2	-2906*	-2923*	3	-2922.5	3.24	427.6933
	3	-2904	-2921	3	-2921.5	2.99	427.8237
	4*	n/a	n/a	n/a	-2922.8	3.72	427.4915*
	5*	-2890	-2919	5	-3036.0*	-110.78	427.6208

2.3.3 Conclusions

The main concern in these results is that the AIC and BIC, the typical methods for parametric model comparison in the literature, only once choose something other than a flat model for $w(z)$, when Ω_m is known and $w(z)$ is highly curved. If $w(z)$ is only lightly curved, BIC and AIC incorrectly choose a flat model. No model comparison methods perform well when Ω_m is unknown because there are multiple solutions issue. We need to address the amount of uncertainty associated with Ω_m before the results for the fit of $w(z)$ are coherent and we plan to do that by adding CMB and BAO data in Chapter 3.

Table 2.4: Model Comparison - Variable Δ and Ω_m

Data	Model	BIC	AIC	p	DIC	p_d	PPL
μ_1	0*	-2907*	-2924*	3	-2924.2	3.01	427.3903
	1	-2899	-2922	4	-2922.6	3.88	427.3608*
	2	-2892	-2920	5	-2926.0	-0.37	427.4834
	3	-2892	-2920	5	-2925.6	0.45	427.4248
	4	n/a	n/a	n/a	-2924.2	2.32	427.4208
	5	-2876	-2916	7	-2927.9*	-1.73	427.4468
μ_2	0	-2899	-2916	3	-2915.8	2.96	428.4060
	1	-2899*	-2922*	4	-2923.2	3.53	427.5772
	2	-2892	-2920	5	-2925.5	0.62	427.3301
	3*	-2892	-2920	5	-2926.6*	-0.3	427.5605
	4*	n/a	n/a	n/a	-2922.4	4.01	427.4432
	5*	-2876	-2916	7	-2922.4	3.66	427.3113*
μ_3	0	-2894	-2911	3	-2912.7	2.39	428.3618
	1	-2899*	-2922*	4	-2922.0	3.76	427.5172
	2	-2891	-2920	5	-2930.2*	-4.42	427.4444*
	3	-2891	-2920	5	-2924.7	0.6	427.6364
	4*	n/a	n/a	n/a	-2928.127	-2.14	427.4649
	5*	-2876	-2916	7	-2928.1	-2.36	427.5131

The DIC provides some useful information about the estimated effective number of parameters. The effective number of parameters do not seem to be very accurate when Ω_m and Δ are variable; we assume this is because of the interdependencies and multiple solutions problem discussed in Section 1.3.4. Looking at Table 2.3 when Ω_m is fixed brings some interesting observations. The parametric models have about the same estimated effective number of parameters as the counted number of parameters. The parameters tend to be correlated and some have informative priors; each parameter may not add as much new information to the model, so the DIC counts may be lower. The interesting thing is that Model 4, the non-parametric GP model, has nearly the same number of effective parameters as the two-parameter parametric forms of $w(z)$.

Overall, the simulated data helps show that the GP is quite competitive with the other parametric models. The effective number of parameters help to show how the GP can be competitive. It offers the advantage of being flexible without over parameterizing the model or increasing the uncertainty estimation of $w(z)$. But the issues with the uncertainty

associated Ω_m and its interaction with $w(z)$ need more work. To constrain the interaction and Ω_m we plan on doing a comprehensive study in Chapter 3 using the BAO and CMB data directly instead of through a prior.

2.4 Real Datasets

Until now we have provided results based on simulated data to show the power of the different models. The simulated data is what is expected to be collected in the near future with $n \approx 2000$ observations. The simulated data verifies the models as the truth is known and also ensures we can computationally handle the larger future datasets. Currently, the real data has fewer points with more noise than the simulated data. For the real data, we show all models discussed thus far and do a full model comparison and hypothesis test for $w = -1$ in this section.

Over the past few years, there has been a series of datasets available for the supernovae Ia (SNe Ia). There are five datasets we examine: the Davis dataset (n=192) (Davis et al., 2007; Riess et al., 2007; Wood-Vasey et al., 2007), Union dataset (n=307) (Kowalski and Rubin, 2008), Constitution dataset (n=397) (Hicken et al., 2009), MLCS17 dataset (n=372), and Union 2 (n=557) (Amanullah et al., 2010). Four of these sets are fit using SALT light curve fitter and one uses the MLCS17 light curve fitter to get the distance measure, μ (Guy et al., 2005). The real data has larger error bars associated with μ and less data points than in our simulated data which produces much larger PI bands for $w(z)$. Figure 2.5 and 2.6 display the distance-redshift relationship and uncertainty measure associated with μ for the real datasets.

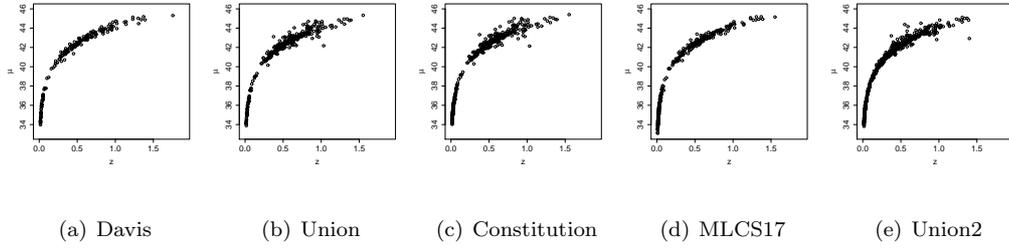


Figure 2.5: z vs. μ for the real SNe datasets

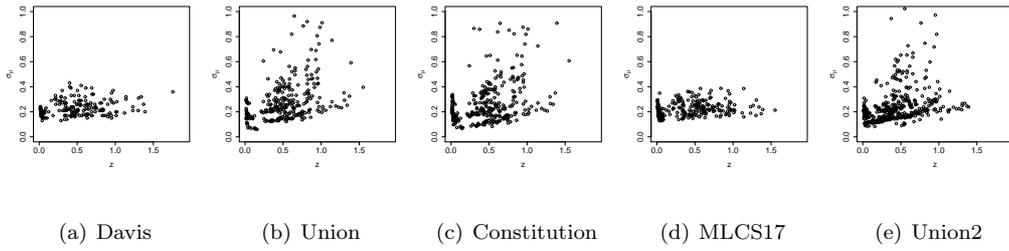


Figure 2.6: Error measures associated with μ for the real SNe datasets

2.4.1 Analysis and Results

We fit the null hypothesis model with $w(z) = -1$ (Model 0) and all three parametric models from Chapter 1: $w(z) = a$, $w(z) = a + bz$, and $w(z) = a + b(\frac{1}{1+z} - 1)$ (Models 1-3). Three non-parametric models from Chapter 2 are also fit; two GP models (Models 4e and 4g) and one damped Hermite basis expansion model (Model 5). The GP was initially run with an exponential correlation function (Model 4e); the results were mostly flat with some slight curvature, as seen in Figure 2.10. Because of these rather flat results, we can use an approximately Gaussian correlation function ($\alpha = 1.9999$) as the second GP model (Model 4g) without concern of over-smoothing.

In this analysis of real data, Ω_m and Δ cannot ever be fixed parameters because unlike the simulated data, we do not know the truth. We give them the same priors as we did in the simulated data. H_0 is set to 70 for the two Union datasets and is 65 for the other three. Δ is an offset parameter which incorporates the uncertainty associated with H_0 . The results of the analysis for all the models is found in Table 2.5 and 2.6 and Figure 2.7, 2.8, 2.9, 2.10, 2.11, and 2.12. The results of the non-parametric GP model has also been published in Holsclaw et al. (2010a).

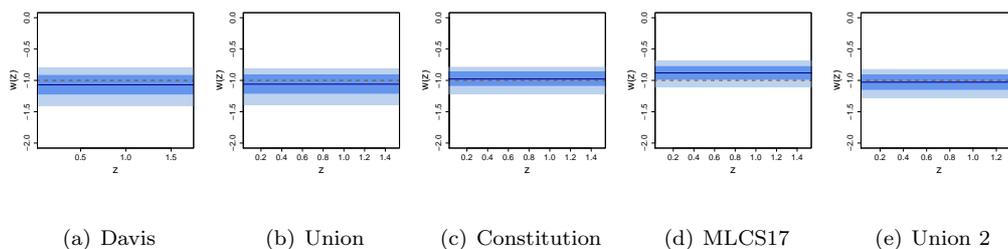


Figure 2.7: Model 1 fits for the real SNe datasets

Table 2.5: All Models - 95% Posterior PIs for Real SNe Data

Model	Data	a	b	Ω_{m0}	Δ	σ^2
0	Davis	n/a	n/a	(0.22,0.33)	(-0.07,0.03)	(0.85,1.27)
	Union	n/a	n/a	(0.24,0.33)	(-0.04,0.03)	(0.88,1.22)
	Const.	n/a	n/a	(0.24,0.33)	(-0.04,0.02)	(1.03,1.36)
	MLCS17	n/a	n/a	(0.27,0.35)	(-0.03,0.02)	(0.94,1.26)
	Union2	n/a	n/a	(0.23,0.31)	(-0.03,0.02)	(0.87,1.10)
1	Davis	(-1.40,-0.80)	n/a	(0.21,0.36)	(-0.09,0.04)	(0.86,1.28)
	Union	(-1.39,-0.81)	n/a	(0.21,0.37)	(-0.05,0.04)	(0.89,1.22)
	Const.	(-1.22,-0.79)	n/a	(0.20,0.35)	(-0.04,0.03)	(1.04,1.37)
	MLCS17	(-1.10,-0.69)	n/a	(0.19,0.35)	(-0.03,0.03)	(0.95,1.26)
	Union2	(-1.28,-0.83)	n/a	(0.21,0.35)	(-0.03,0.02)	(0.87,1.11)
2	Davis	(-1.71,-0.55)	(-3.33,2.12)	(0.21,0.37)	(-0.10,0.04)	(0.85,1.28)
	Union	(-1.62,-0.76)	(-2.30,2.34)	(0.21,0.38)	(-0.07,0.03)	(0.88,1.21)
	Const.	(-1.31,-0.42)	(-4.78,1.37)	(0.20,0.39)	(-0.04,0.04)	(1.03,1.37)
	MLCS17	(-1.20,-0.36)	(-3.43,1.22)	(0.20,0.37)	(-0.03,0.04)	(0.94,1.26)
	Union2	(-1.30,-0.72)	(-2.68,1.08)	(0.21,0.37)	(-0.04,0.03)	(0.88,1.10)
3	Davis	(-1.76,-0.48)	(-3.53,4.71)	(0.21,0.37)	(-0.10,0.04)	(0.86,1.28)
	Union	(-1.73,-0.71)	(-3.71,3.19)	(0.21,0.38)	(-0.08,0.03)	(0.88,1.21)
	Const.	(-1.29,-0.05)	(-2.04,9.51)	(0.22,0.41)	(-0.04,0.05)	(1.03,1.36)
	MLCS17	(-1.25,-0.34)	(-1.98,4.19)	(0.21,0.36)	(-0.03,0.04)	(0.94,1.26)
	Union2	(-1.35,-0.65)	(-1.56,4.36)	(0.21,0.38)	(-0.03,0.03)	(0.87,1.11)
4e	Davis	n/a	n/a	(0.21,0.35)	(-0.09,0.03)	(0.85,1.27)
	Union	n/a	n/a	(0.22,0.36)	(-0.06,0.04)	0.88,1.21)
	Const.	n/a	n/a	(0.20,0.35)	(-0.04,0.03)	(1.03,1.36)
	MLCS17	n/a	n/a	(0.20,0.35)	(-0.03,0.03)	(0.94,1.26)
	Union2	n/a	n/a	(0.21,0.35)	(-0.03,0.02)	(0.88,1.11)
4g	Davis	n/a	n/a	(0.21,0.35)	(-0.08,0.04)	(0.85,1.28)
	Union	n/a	n/a	(0.21,0.36)	(-0.05,0.04)	(0.88,1.21)
	Const.	n/a	n/a	(0.21,0.35)	(-0.04,0.03)	(1.03,1.37)
	MLCS17	n/a	n/a	(0.20,0.35)	(-0.03,0.03)	(0.94,1.26)
	Union2	n/a	n/a	(0.21,0.34)	(-0.03,0.02)	(0.87,1.11)
5	Davis	(-0.43,0.27)	(-0.71,0.61)	(0.21,0.36)	(-0.09,0.04)	(0.85,1.27)
	Union	(-0.45,0.18)	(-0.46,0.78)	(0.21,0.36)	(-0.06,0.03)	(0.88,1.21)
	Const.	(-0.26,0.28)	(-0.77,0.50)	(0.21,0.36)	(-0.04,0.03)	(1.03,1.36)
	MLCS17	(-0.17,0.39)	(-0.71,0.45)	(0.20,0.36)	(-0.03,0.03)	(0.94,1.26)
	Union2	(-0.28,0.19)	(-0.65,0.43)	(0.20,0.36)	(-0.03,0.03)	(0.87,1.11)

Table 2.6: Model Comparison for the Real Data

Data	Model	BIC	AIC	p	DIC	p_d	PPL
Davis	0	-11.5*	-21.3*	3	-21.371*	2.95	60.198
	1	-6.5	-19.5	4	-21.269	3.00	59.899
	2	-1.3	-17.6	5	-19.987	3.52	59.795*
	3	-1.2	-17.5	5	-19.897	3.61	59.876
	4e	n/a	n/a	n/a	-21.234	3.01	59.967
	4g	n/a	n/a	n/a	-21.332	2.95	60.011
	5	14.4	-11.6	8	-21.001	3.06	59.919
Union	0	31.0*	19.8*	3	19.973	3.04	128.092
	1	35.6	20.6	4	19.666	2.90	128.006
	2	41.2	22.5	5	20.182	3.52	127.710*
	3	41.3	22.7	5	20.196	3.46	127.954
	4e	n/a	n/a	n/a	19.687	3.01	128.008
	4g	n/a	n/a	n/a	19.615*	2.97	127.920
	5	58.4	28.5	8	20.113	3.61	127.892
Const.	0	4.0*	-7.9*	3	-7.489	3.19	158.622
	1	10.0	-5.9	4	-7.837*	2.98	158.286*
	2	12.3	-7.6	5	-7.531	2.92	158.432
	3	12.8	-7.1	5	-7.442	3.04	158.365
	4e	n/a	n/a	n/a	-7.604	3.22	158.316
	4g	n/a	n/a	n/a	-7.510	3.18	158.342
	5	32.4	0.5	8	-7.325	3.16	158.326
MLCS17	0	-78.3*	-90.1*	3	-90.354*	2.74	109.436
	1	-72.7	-88.3	4	-90.208	3.01	108.989
	2	-66.9	-86.5	5	-88.189	3.86	109.079
	3	-66.9	-86.5	5	-88.775	3.67	108.977*
	4e	n/a	n/a	n/a	-89.826	3.24	109.041
	4g	n/a	n/a	n/a	-89.938	3.18	109.120
	5	-49.2	-80.6	8	-89.281	3.48	109.113
Union2	0	-201.3*	-214.3*	3	-213.474	3.40	183.869
	1	-195.1	-212.3	4	-213.990*	3.12	183.848
	2	-189.3	-210.9	5	-212.882	3.56	183.736
	3	-189.7	-211.3	5	-212.865	3.64	183.653*
	4e	n/a	n/a	n/a	-213.754	3.28	183.833
	4g	n/a	n/a	n/a	-213.808	3.21	183.807
	5	-170.4	-204.9	8	-213.436	3.28	183.784

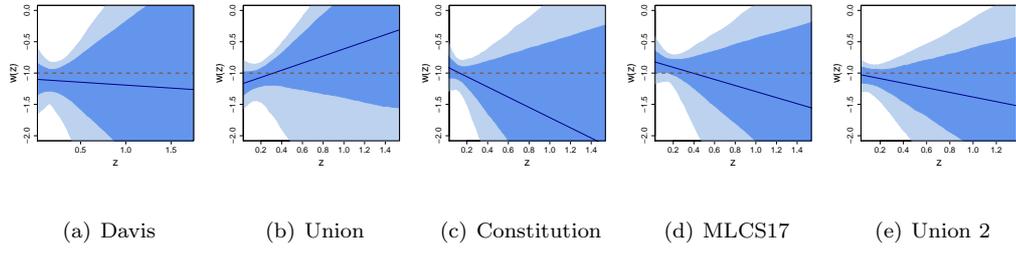


Figure 2.8: Model 2 fits for the real SNe datasets

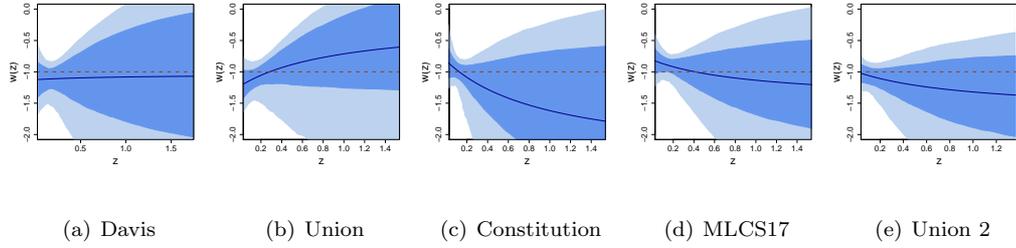


Figure 2.9: Model 3 fits for the real SNe datasets

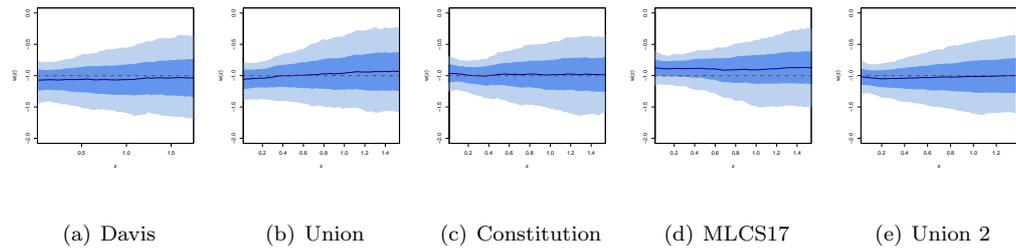


Figure 2.10: Model 4 - GP ($\alpha = 1$) fits for $w(z)$ using the real SNe data

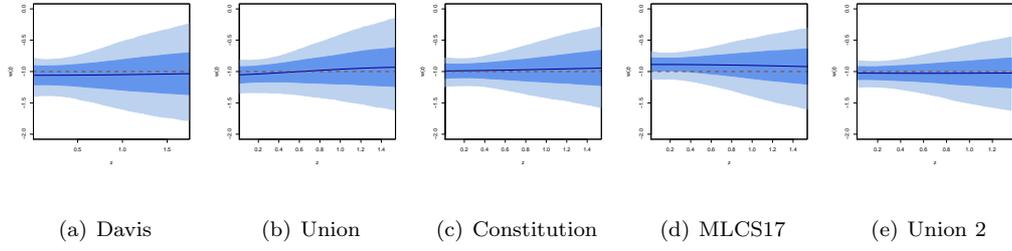


Figure 2.11: Model 4- GP ($\alpha = 1.9999$) fits for $w(z)$ using the real SNe data

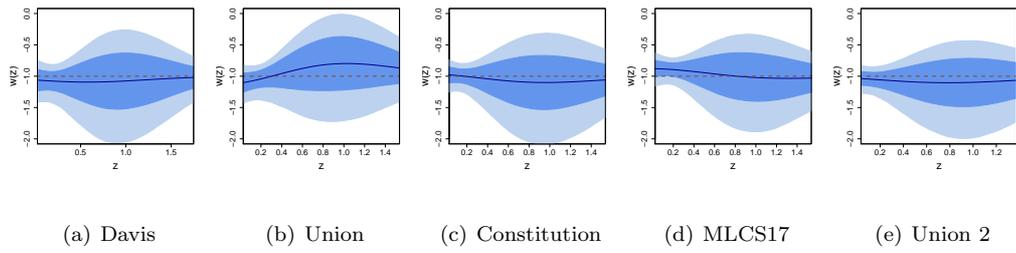


Figure 2.12: Model 5- Hermite basis of polynomials fits for the real SNe datasets

2.4.2 Conclusions

The mixing in the algorithm is good and has stable MCMC posteriors. Every model and dataset is within the 95% PI of having the truth be $w(z) = -1$. This is partly due to the fact that the 95% PI for most of these models are quite large. Better data quality or quantity would help to reduce these bands. We need less uncertainty about the cosmological parameters Δ and Ω_m to better constrain $w(z)$. We gain little to no information about Ω_m through this analysis; the informative prior based on BAO and CMB data is quite similar to the posterior distribution of Ω_m . Ω_m cannot be constrained well by supernova data alone; we need other data sources to shrink its uncertainty and thus the uncertainty associated with dark energy EOS, $w(z)$. We do not comment on the results of Δ because its meaning is not fully defined as it is a mixture of H_0 and the marginalization during light curve fitting process which results in an offset parameter M .

Overall, the non-parametric GP fit of $w(z)$ has tighter PIs than some of the parametric models, specifically the two parameter models which is also the case in the simulated data. The GP analysis have $w(z) = -1$ within the 95% probability but more or better data is needed to reduce uncertainty or more precise data to draw any firm conclusions about $w(z) = -1$.

Chapter 3

Incorporating CMB and BAO

Data

How does data from other probes constrain Ω_m and the dark energy EOS, $w(z)$? Can we expect these other probes to decouple these parameters? Previously, we discussed that Ω_m and $w(z)$ have interdependencies that result in multiple solutions. Therefore, these parameters cannot be constrained by SNe data alone. We were using an informative prior for Ω_m resulting from analysis of other datasets like baryon acoustic oscillation data (BAO) and cosmic microwave background data (CMB) in the previous chapters: $\pi(\Omega_m) \sim N(0.27, .04^2)$ (Hojjati and Pogosian, 2010; Feng et al., 2008). But we wish to include these two probes directly to the analysis and have a flat prior: $\pi(\Omega_m) \sim U(0,1)$. By including the two probes that have different distance-redshift relationships directly to the analysis it should help constrain the multiple solutions problem. In the previous chapter, we show that the GP model is the most flexible option while not increasing uncertainty. In this chapter, we solely focus on the GP model and include more data types to the analysis. Others have done a joint analysis of SNe, BAO, and CMB for the parametric Model 1 and Model 3

(Komatsu et al., 2010; Escamilla-Rivera et al., 2011). There is a concern that the two parameter models presented in Chapter 1 fail to be universal enough to provide coherent fit for the z domain out to infinity. The CMB is measured at high z values and the BAO distance-redshift relationship requires an integration with limits on the z domain of infinity.

A joint analysis of SNe, BAO, and CMB probes which each have a different distance-redshift relation can help constrain the unknowns $w(z)$ and Ω_m . The BAO data is measured at low redshift (z) values similar to that of the supernova data. SNe and BAO data have many observations between $z \in (0, 2)$, which makes for a coherent GP model on this range. The single CMB observation is at high redshift ($z = 1000$) which is problematic for any non-parametric model because that observation is away from the rest. The parametric models assume a rigid form for the entire range $(0, \infty)$ (parametric models need to be carefully examined to ensure their form is theoretically appropriate out to infinity and that they correctly handle the single point at $z = 1000$). We must make some assumptions at this point for our model on the range $z = (2, \infty)$ where only one observation at $z = 1000$ is available and because the GP model does not support an integration over (z^*, ∞) mathematically. We assume the dark energy EOS, $w(z)$, is constant on the range $z = (2, \infty)$ because of the lack of data. Thankfully, the real interest for the time varying form of $w(z)$ is on the range $z = (0, 2)$ and we use a flexible GP model for this range. This model assumes the dark energy EOS is continuous on the entire range.

SNe data is only available for low redshift values, so we have been disregarding the radiation term in the distance-redshift relation. In contrast, the CMB observation is at high redshift and BAO requires integration up to infinity on the redshift range. The radiation term ($\Omega_r(1+z)^4$) in the $G(z)$ equation is now quite important; we no longer consider $\Omega_r(z)$ to be zero in this chapter. $\Omega_r(z)$ is the radiation density for photons and neutrinos. We use

the following formula for $\Omega_r(z)$ when CMB or BAO is added to the analysis:

$$\begin{aligned}\Omega_r(z) &= \Omega_{0r} (1 + 0.2271 N_{eff} f(m_\nu a / T_{\nu 0})) \\ &= \Omega_{0r} \left(1 + 0.2271 N_{eff} \left(1 + \left(0.3173 \left(\frac{187}{1+z} \left(\frac{\Omega_{0r} h^2}{10^{-3}} \right) \right) \right)^{1.83} \right)^{1/1.83} \right)\end{aligned}$$

where $\Omega_{0r} = 4.982 * 10^{-5}$ (from WMAP-7), $h = H_0/100$ Hubble's constant, and $N_{eff} = 3.04$ the standard three neutrino species (Komatsu et al., 2010).

We are assuming the three data sources SNe, CMB, and BAO to be independent of each other, so we can create a separable joint likelihood. SNe data has a variable variance parameter (σ^2) associated with its measures. If we have multiple observations arising from BAO or CMB, we need to assign them a variable variance parameter as well. Here CMB has only a single observation so a variance parameter is not necessary, but the simulated BAO dataset has 20 observations and thus we assign σ_B^2 as its variable variance parameter.

3.1 Simulated Data

We continue to use the same simulated data for the SNe ($n \approx 2000$), which are observations that are planned to be obtain in the near future (this data was also used in Chapter 1 and 2 and (Holsclaw et al., 2010b)). Now we include a single simulated CMB point which is very similar to what is available and twenty simulated BAO data observations which are similar to data which should be available in the near future. In reality, there is only one CMB point, two BAO points, and about 557 SNe (Union 2 dataset) observations currently available to us (see Section 3.5 for a discussion on the currently available data (Amanullah et al., 2010)). The simulated data is created with the same parameters as the supernova data ($\Delta = 0$, $\Omega_m = 0.27$, $\sigma^2 = 1$, σ_B^2 and $H_0 = 70.4$). But the distance-redshift relation is different for each probe.

3.2 Cosmic Microwave Background Data

First, we describe the process of adding the one CMB observation to the existing SNe analysis. The cosmic microwave background data (CMB) comes from the Wilkinson Microwave Anisotropy Probe (WMAP) (Bennett et al., 2003a,b). The physics of this data is best understood compared to the other two probes and should shrink the uncertainty of the estimated parameters. This probe explores very high redshift, many orders of magnitude higher than any of the other data types discussed. There is a single data point available for this probe near redshift $z = 1000$ (we refer to this point as z^*). We include this point by assuming it is independent to the SNe dataset and using a joint likelihood.

The simulated CMB data point is created with the same parameters as the supernova data ($\Delta = 0$, $\Omega_m = 0.27$, $\sigma^2 = 1$ and $H_0 = 70.4$) but using the CMB equations relating distance (y^*) and redshift (z^*). We assume that $y^* \sim N(R(z^*), \tau^{*2})$, where $R(z^*)$ is the transform function given in equation (3.1) and (3.2) and $\tau_{z^*}^2$ is the measurement error associated with the observation y^* (Wang and Mukherjee, 2007; Elgaroy and Multamaki, 2007; Corasaniti and Melchiorri, 2008; Bond et al., 1997). $z^* = 1090.89$ and true $R(z^*) = (1.7226, 1.7024, 1.6699)$ (respectively for μ_1 , μ_2 and μ_3) and $\tau^* = 0.019$. The simulated values used for the analysis are $y^* = (1.7361, 1.7158, 1.6834)$. The relation between this form of the data is given by:

$$R(z^*) = \sqrt{\Omega_m} \int_0^{z^*} G(s) ds \quad (3.1)$$

$$G(s) = \left(\Omega_r(s)(1+s)^4 + \Omega_m(1+s)^3 + (1 - \Omega_r(s) - \Omega_m)(1+s)^3 e^{-3 \int_0^s \frac{-w(u)}{1+u} du} \right)^{-1/2} \quad (3.2)$$

The SNe and CMB data have a joint likelihood taking the form:

$$L \propto \sigma^{-n} \exp -\frac{1}{2} \sum_{i=1}^n \left(\frac{\mu_i - T(z)}{\tau_i \sigma} \right)^2 \exp -\frac{1}{2} \left(\frac{y^* - R(z^*)}{\tau^*} \right)^2$$

H_0 and Ω_{0r} are fixed parameters. We have unknown cosmological parameters: Ω_m , Δ , and σ^2 ; which all have flat or non-informative priors in this analysis. Ω_m no longer has an

informative prior based on the CMB and BAO data because that data is being incorporated directly. The GP has covariance parameters, κ^2 and ρ , that have the same priors from previous analysis of SNe data.

3.3 Baryon Acoustic Oscillation Data

The second probe of interest to add to the SNe analysis is baryon acoustic oscillation (BAO) data that comes from measurements of clustered baryonic matter or large scale structure of matter like galaxies. This probe provides a standard ruler for distance measures that are used in a distance-redshift relationship.

We simulate twenty BAO data points which are similar to what surveys like BigBOSS hope to obtain in the near future (Schlegel et al., 2009b,a). Each BAO point has a redshift value and two associated observed distance measures with estimated error structure generated from Seo and Eisenstein (2007). We assume that each distance observation for the BAO observations (y_{1i} and y_{2i}) have correlated bivariate Normal distribution as seen in equation (3.3) where we assume $\Omega_{0b}/\Omega_r = 915.35$ and $z_d = 1020.5$; H_0 cancels out of these equations.

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left[\begin{bmatrix} D_A(z_i)/r_s \\ H(z_i)r_s \end{bmatrix}, \sigma_B^2 K \right] \quad \text{where} \quad K = \begin{bmatrix} \sigma_{y_{1i}}^2 & r_{12i}\sigma_{y_{1i}}\sigma_{y_{2i}} \\ r_{21i}\sigma_{y_{1i}}\sigma_{y_{2i}} & \sigma_{y_{2i}}^2 \end{bmatrix} \quad (3.3)$$

$$H(z) = H_0/G(z) \quad (3.4)$$

$$D_A(z) = \frac{c}{1+z} \int_0^z \frac{G(s)}{H_0} ds \quad (3.5)$$

$$r_s = \frac{c}{\sqrt{(3)}} \int_{z_d}^{\infty} G(s) \left(H_0 \sqrt{1 + \frac{3\Omega_{ob}}{4\Omega_r(1+z)}} \right)^{-1} ds \quad (3.6)$$

The simulated dataset has twenty BAO points ranging from redshift $[0, 2]$. The bi-variate likelihood for the BAO data is:

$$L \propto \left(\frac{1}{\sigma_B}\right)^{2m} \exp \left[-\frac{1}{2\sigma_B^2} \sum_{j=1}^m \left(\begin{pmatrix} y_{1j} - \frac{D_A(z)/r_s}{H(z)r_s} \\ y_{2j} - \frac{D_A(z)/r_s}{H(z)r_s} \end{pmatrix}' K_j^{-1} \begin{pmatrix} y_{1j} - \frac{D_A(z)/r_s}{H(z)r_s} \\ y_{2j} - \frac{D_A(z)/r_s}{H(z)r_s} \end{pmatrix} \right) \right]$$

As with the CMB data: H_0 , Ω_r , and Ω_{0b} are fixed parameters and we have unknown parameters: Ω_m , Δ , σ^2 , and σ_B^2 with either flat or non-informative priors. The GP has covariance parameters κ^2 and ρ which have somewhat informative priors described in the SNe analysis.

3.4 Results for Combined Data Sources

The SNe can be combined with BAO, CMB, or both BAO and CMB data in a joint likelihood; we assume that the different probes are independent. The results of this analysis are shown in Figures 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6 and Table 3.1. The table does not include σ^2 or σ_B^2 as they are not interesting parameters. σ^2 has a 95% PI of (0.92,1.03) for all datasets.

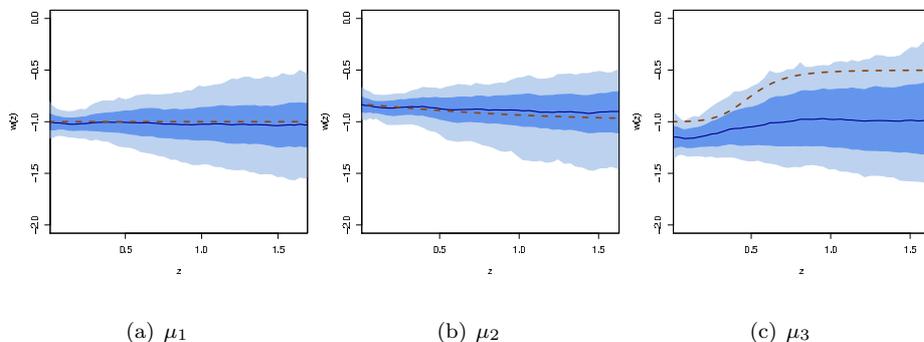


Figure 3.1: GP Model - SNe

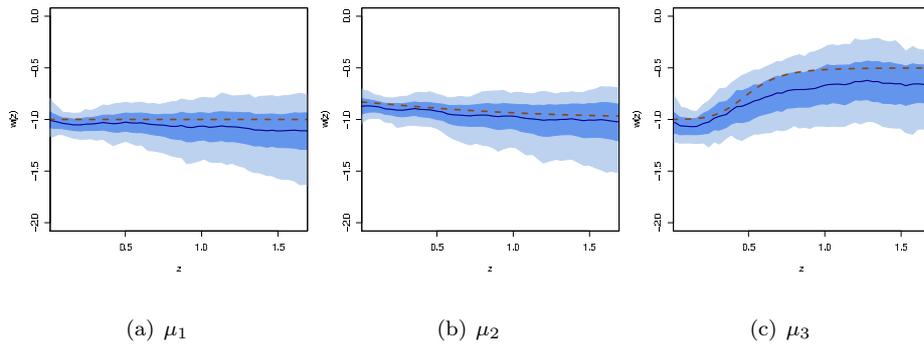


Figure 3.2: GP Model - SNe+CMB

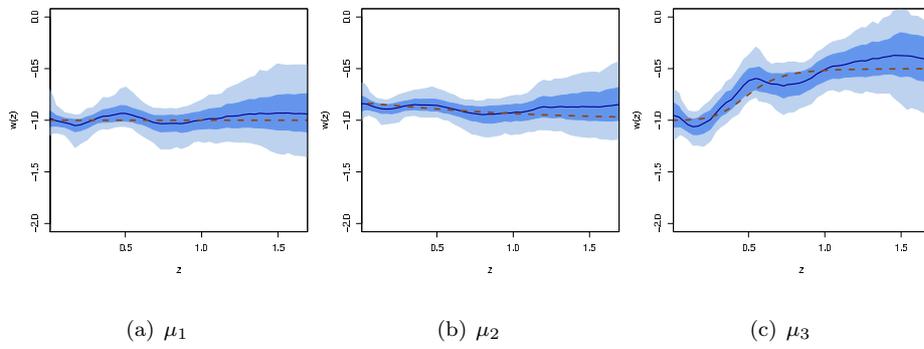


Figure 3.3: GP Model - SNe+BAO

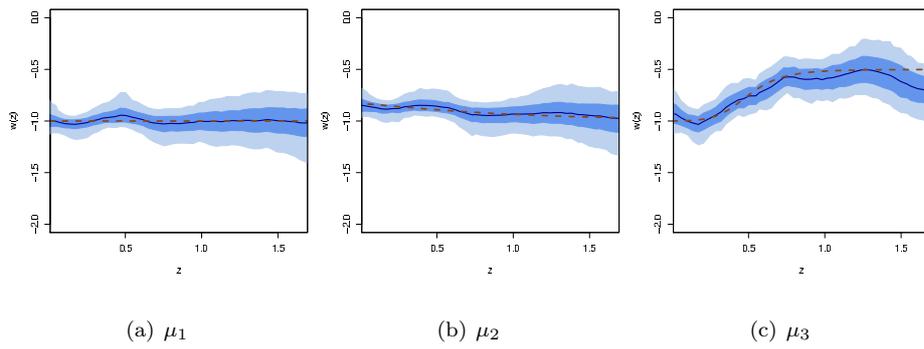


Figure 3.4: GP Model - SNe+CMB+BAO

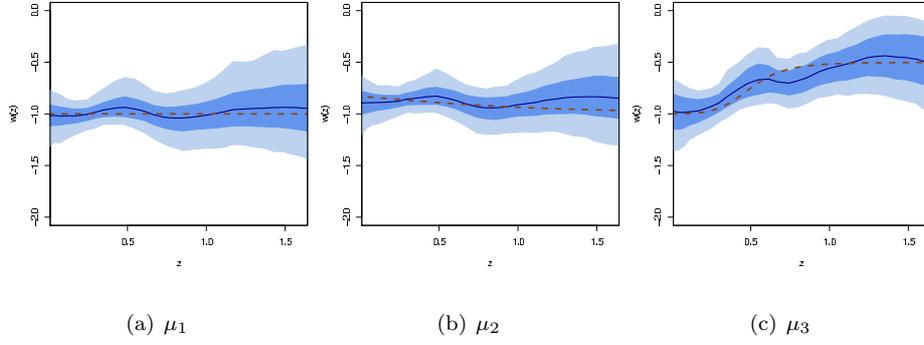


Figure 3.5: GP Model - BAO

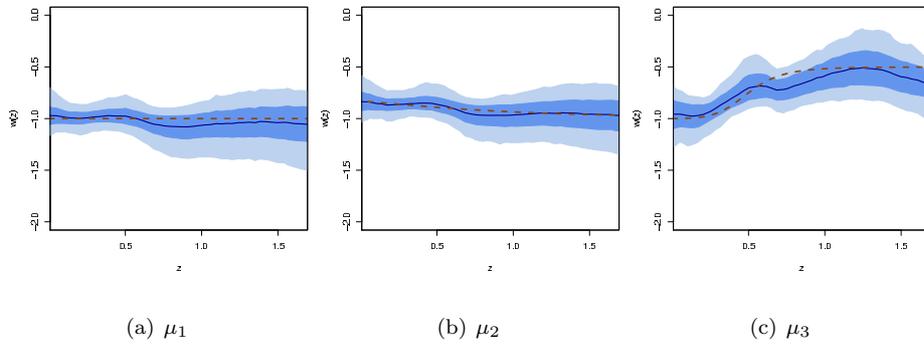


Figure 3.6: GP Model - BAO+CMB

Table 3.1: Model 4 - Simulated Data - Posterior 95% PIs

Data Type	Data	Ω_m	Δ	ρ	κ^2
SNe	μ_1	(0.23,0.30)	(-0.02,0.02)	(0.631,0.998)	(0.16,0.72)
	μ_2	(0.21,0.31)	(-0.02,0.01)	(0.630,0.998)	(0.16,0.74)
	μ_3	(0.26,0.37)	(-0.03,0.01)	(0.532,0.995)	(0.16,0.75)
SNe+CMB	μ_1	(0.25,0.30)	(-0.02,0.02)	(0.568,0.997)	(0.16,0.73)
	μ_2	(0.25,0.31)	(-0.02,0.01)	(0.605,0.999)	(0.16,0.72)
	μ_3	(0.25,0.34)	(-0.02,0.02)	(0.521,0.972)	(0.16,0.88)
SNe+BAO	μ_1	(0.26,0.28)	(-0.02,0.02)	(0.510,0.998)	(0.16,0.80)
	μ_2	(0.26,0.28)	(-0.02,0.02)	(0.612,0.998)	(0.16,0.74)
	μ_3^*	(0.24,0.28)	(-0.02,0.02)	(0.454,0.934)	(0.19,1.04)
SNe+BAO+CMB	μ_1	(0.26,0.28)	(-0.02,0.02)	(0.596,0.999)	(0.16,0.77)
	μ_2	(0.26,0.28)	(-0.02,0.01)	(0.620,0.998)	(0.15,0.74)
	μ_3	(0.26,0.28)	(-0.02,0.02)	(0.452,0.946)	(0.17,0.87)
BAO	μ_1	(0.22,0.31)	n/a	(0.549,0.997)	(0.16,0.76)
	μ_2	(0.22,0.30)	n/a	(0.556,0.997)	(0.16,0.76)
	μ_3	(0.22,0.33)	n/a	(0.468,0.967)	(0.17,0.82)
BAO+CMB	μ_1	(0.25,0.31)	n/a	(0.576,0.996)	(0.16,0.76)
	μ_2	(0.25,0.31)	n/a	(0.564,0.996)	(0.16,0.79)
	μ_3	(0.24,0.32)	n/a	(0.458,0.976)	(0.17,0.87)

In the BAO and BAO+CMB analysis, the fit is a bit wavy but this is due to only using twenty points unlike the $n \approx 2000$ SNe points. We only show one realization of generated errors in this paper, which also accounts for the waviness. Each realization of errors would produce a different waviness.

The best fit seems to result from a full combination of SNe, BAO, and CMB. But adding either BAO or CMB to the SNe data also is an improvement in fit and also in estimation of Ω_m . This is not necessarily expected. In the previous chapters, the prior on Ω_m is informed by CMB and BAO data that was not included in the analysis. Here we add the CMB and BAO data directly to the likelihood equation and amend the prior on Ω_m to be flat and non-informative. It seems that these two different analysis should yield nearly the same results but we find that adding either of these data sources directly in with the SNe data is quite an improvement in most cases over having an informative prior on Ω_m . One thought is that the informative prior on Ω_m is too lax as $\pi(\Omega_m) \sim N(0.27, 0.04^2)$ and it should have smaller standard error. But another thought is that the other probes have

different distance-redshift relationships and thus different correlation on the Ω_m and $w(z)$ space, which may reduce the parameter interdependency issues discussed in Section 1.3.4. Overall, it is best to include all of the data sources directly to the analysis.

3.5 Real Data Analysis

Currently we have available 557 SNe observations ($z = (0.015, 1.4)$), 2 BAO data points ($z=0.20,0.35$), and one CMB data point ($z= 1090.79$) (Percival et al., 2010; Bond et al., 1997). We assume a GP model with exponential correlation function on $z = (0, 1.4]$ and $w(z) = \text{const}$ on $z = (1.4, \infty)$. The same analysis is performed just as before with the simulated data but now with a bit of a change to the BAO distance-redshift equation. Currently, the only measurement for BAO is in terms of: $r_s(z_d)/D_V(z)$ where $D_V(z) = (cz(1+z)^2 D_A^2(z)/H(z))^{1/3}$, whereas in the simulated data there are two correlated distance measures for each observation.

The real data has observed redshift values and also distance measures with stated error bars: for CMB $z = 1090.79$ with $R = 1.719_{-0.019}^{+0.019}$ and for BAO $z = 0.20$ with $r_s(z_d)/D_V(z) = 0.1905_{-0.0061}^{+0.0061}$ and $z = 0.35$ with $r_s(z_d)/D_V(z) = 0.1097_{-0.0036}^{+0.0036}$ (Percival et al., 2010). A few of the parameters have slightly different values with the real data than in the simulated sets: $\Omega_r = 0.00004897$, $H_0 = 70$, $z_d = 1020.3$, and $\Omega_{0b}/\Omega_r = 914.54$. The priors for Δ , Ω_m , σ^2 , κ , and ρ remain the same.

Table 3.2: Model 4 - Union 2 SNe Dataset - Posterior 95% PIs

Data Type	Ω_m	Δ	σ_2	ρ	κ^2
SNe	(0.21,0.35)	(-0.03,0.03)	(0.88,1.11)	(0.567,0.996)	(0.16,0.75)
SNe+BAO	(0.25,0.35)	(-0.03,0.02)	(0.87,1.10)	(0.541,0.996)	(0.16,0.80)
SNe+CMB	(0.23,0.32)	(-0.03,0.03)	(0.87,1.10)	(0.532,0.996)	(0.16,0.81)
SNe+BAO+CMB	(0.25,0.33)	(-0.03,0.02)	(0.87,1.10)	(0.567,0.996)	(0.17,0.80)

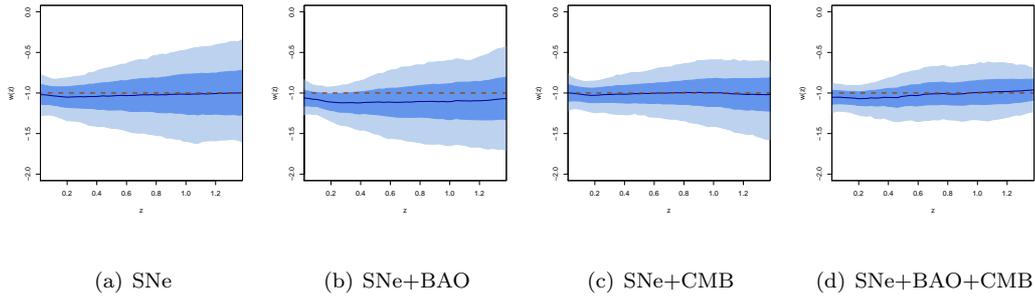


Figure 3.7: Real Data - Union 2 SNe data and BAO and CMB data points

Better data quality or more observations are needed for a more certain conclusion to be made about $w(z)$ and the other cosmological parameters. We present the current data results here and quickly move on to a discussion in Chapter 4 about what data is most needed to constrain the dark energy EOS.

Chapter 4

Experimental Design

Another way $w(z)$ can be further constrained is by improving the quantity and quality of the supernova or BAO data. Much of the uncertainty reduction is up to the astronomers, as they work to improve current telescope resolution and shrink the noise associated with their measures. But statistically, we should be able to answer a foundational question that bridges the theoretical cosmologist and observational astronomer as to where more data is needed on the redshift range. This chapter address where more SNe data should be collected on the z (redshift) axis to best constrain $w(z)$. Typically, uncertainty is lessened with more observations. These observations are expensive to collect so it is necessary to target regions. This is not only important to the cosmologist and astronomer, but also to the decision maker who must allocate resources for new telescopes or purchase time on existing telescopes.

4.1 Cluster Analysis

The real SNe data that has been collected comes from different telescope surveys, which tend to collect observations within a certain redshift range. Ground based telescopes

take measures that are at lower redshift and satellite telescopes like Hubble take measurements that are at higher redshift. We want to see if any clustering methods could pick up on the differences between these surveys. This may help with the experimental design question of where additional data is needed or which telescope survey is providing the best information at constraining $w(z)$.

We first take a dataset, where we have a pretty good idea of the truth of which SNe come from which telescope survey; namely the Constitution dataset is well documented in Hicken et al. (2009). SNe have naming conventions resulting from which telescope survey they come from. This is how we classify the truth and then use cluster analysis to see if we can recreate the information about which SNe come from which telescope survey.

We have the several output variables from the light curve fitter: z , μ , τ which is the weight associated with μ , and the year the SNe exploded. Figure 4.1 shows the two way interaction plots for these four variables. We color each of the four telescope surveys. The light blue is actually a mix of several ground based surveys; we see they are at low z value. In earlier years, only ground based observations are available. SNe observations are released slowly, some of the more recent years may not have all of the data from all of the telescopes documented yet.

We ran a number of clustering algorithms to discover if any would be able to find the telescope groupings within the dataset. These types of clustering algorithms are routinely used for classification. They should partition SNe observations into groupings by their similarities. All of these clustering algorithms can be found in the `mclust` library of R (Fraley and Raftery, 2006). We assume that we know there are four clusters to make for more precision in the algorithms. In general, none of these algorithms need to know the number of clusters unless we state otherwise.

One type of clustering algorithm is the agglomerative hierarchical clustering (`hc` in R) algorithm (Tan et al., 2006). First, it is hierarchical meaning that it knows about

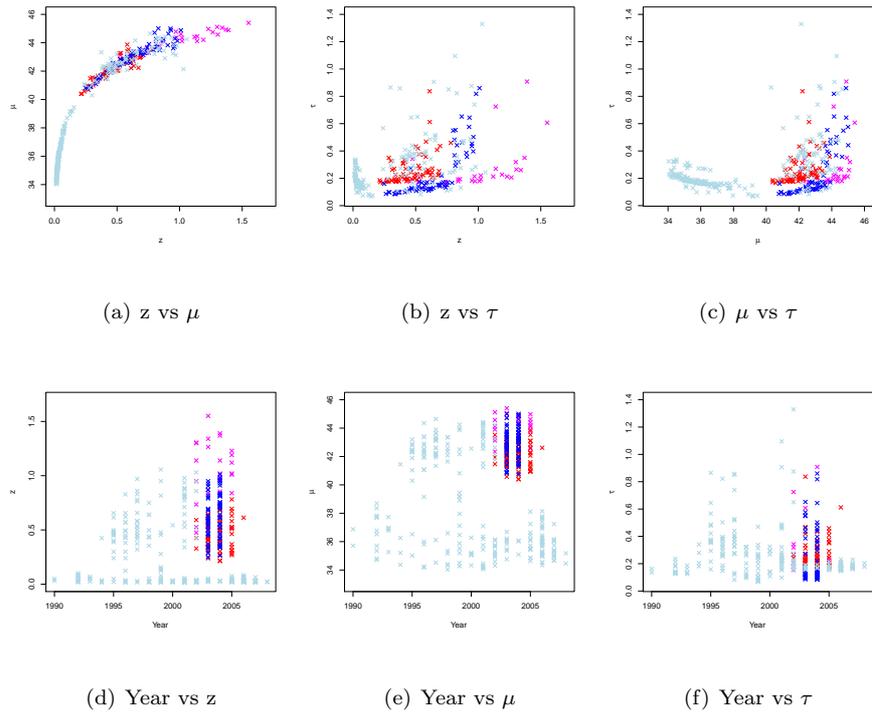


Figure 4.1: Two-way plots of the raw data colored by the four telescope surveys

the sub-clusters as opposed to being just a partitioning algorithm that divides the data into distinct non-overlapping sets. The hierarchical method has subclusters and views the cluster structure as a tree plot called a dendrogram. The agglomerative part of the algorithm refers to the fact that it considers each data observation as its own cluster and then begins grouping them together into larger clusters. This is just the opposite from divisive hierarchical algorithm which consider the data as one cluster and then create sub-cluster structure from there. This particular algorithm is based on Gaussian mixture models using maximum likelihood methods and eigenvalue decomposition (Banfield and Raftery, 1993).

The second broad type of clustering algorithms are also agglomerative hierarchical tree structures (`hclust` in R). This type of cluster analysis relies on dissimilarity matrices or either euclidean or maximum distances. There are a number of methods that fit this general class; here we examine the results from six of these algorithms: single linkage (MIN), complete linkage (MAX), group averages, group medians, centroid, and Ward's minimum variance method. These terms mostly speak of the way two clusters are related to one another (Tan et al., 2006). The single linkage (MIN) algorithm looks at all current clusters of the tree (we begin with each observation being its own cluster at the first level of the tree) and relates the two closest points in different clusters. These two clusters will be merged in the next stage. Complete linkage looks at the furthest points in different clusters and merges the clusters with the smallest maximum distance. Group averages classification looks at all links of the observations in one cluster with another cluster and computes an average. The two clusters with smallest average will be merged. Group median is similar but uses the group median instead of averages. Clusters can be defined by their centroid and these can be compared in the centroid algorithm. Ward's method minimizes the distance of points to their centroid that happens if the two clusters are merged.

K-means clustering is a partitioning algorithm rather than a hierarchical one (`kmeans` in R). The K-means method of clustering relies on the centroid of the clusters.

It tries to minimize the distance each point is to the center of its cluster (Tan et al., 2006). The number of clusters is set initially and those centroids are set initially and then moved about until the algorithm has a best fit.

The last type of clustering we use is model based clustering (mclust in R). This fits many Gaussian mixture models to the data using an EM algorithm and maximum-likelihood principles (Fraley and Raftery, 2006). Then the optimal model is chosen according to the BIC model comparison test. This model is best at finding spheres and ellipses of all shapes and sizes that are Gaussian in nature.

Table 4.1: Success Rate of Classification

Clustering Algorithm	z and μ	plus τ	plus year
Agglomerative hier.	0.607	0.547	0.549
Single linkage	0.572	0.572	0.572
Complete linkage	0.647	0.569	0.587
Group averages	0.652	0.572	0.574
Group medians	0.572	0.574	0.572
Centroid	0.652	0.572	0.572
Ward's method	0.632	0.642	0.524
K-means	0.594	0.642	0.630
Model based	0.582	0.554	0.531

Now that we have laid out some of the different cluster algorithms, we do data analysis. We analyze just z and μ first (results displayed in Table 4.1. The results are a simple proportion of how many of the points are in the correct clusters. Most of the algorithms seem to cluster about 60% correctly. We add two other variables, τ and Year, to the analysis and find they do not help the clustering algorithm to distinguish between telescopes. We also have plots of the cluster analysis in Figure 4.2. Figures 4.2(a) and 4.2(b) contain the truth. None of these methods seem to perform well and some of the methods mainly group all of the points in one large cluster.

We find that the telescope survey does not influence the analysis enough to be needed in the general models. Furthermore, we could not get finalized data about which telescope each SNe came from and it also seems that some SNe are measured through a

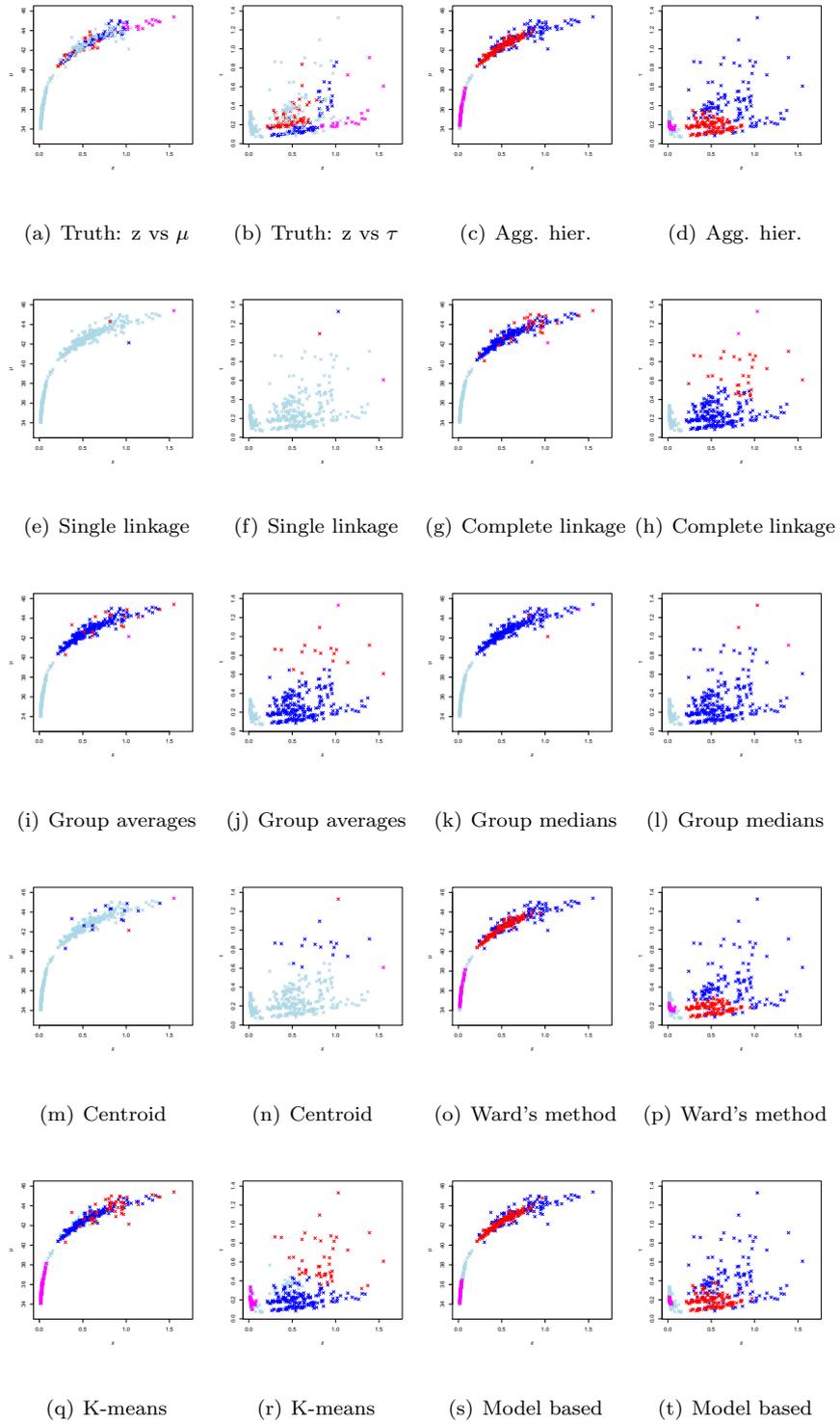


Figure 4.2: Some two-way plots colored by the four telescope surveys when clustering algorithms are run for z , μ , τ , and Year. The first two plots are of the truth. Each of the clustering methods is then displayed in two plots.

mixture of telescopes in some cases. But these results were still of interest to the cosmologists as another way of viewing the data. And the cluster analysis does not answer where more data is needed on the redshift range or which telescope survey may be providing the “best” data for constraining $w(z)$.

4.2 Future Telescope Surveys

Currently, the real data is comprised of several hundred supernova observations. Each of the telescopes search different depths of space for the supernovae and collect their results. We explore which redshift ranges give the best information for constraining $w(z)$. This analysis of redshift range needs to account for any variation in observations as further objects tend to have larger error bars. Ultimately, we would like to answer which of the proposed surveys is best. These lines of investigation are relevant, as many observers continue to collect this supernova data and telescopes range widely in cost.

This analysis may help guide the observers to collect the most relevant data that would be most efficient to the current dark energy research. This type of problem is referred to as experimental design in statistics. Typical experimental design would specify exact redshifts to collect data but the Universe probably would not comply with providing a supernova at that exact redshift location, so instead we alter the standard approach and allow for a distribution of redshift values as specified by proposed telescope surveys.

As in the previous work, we expect this analysis to be computationally intensive due to the highly non-linear relationship in this problem. Currently, the method that has been used to weight different regions of the redshift range is principal component analysis (Huterer and Starkman, 2003). This requires that $w(z)$ be a piecewise discontinuous function, which is not coherent with the physics of $w(z)$, and the supernovae are binned according to redshift to perform this analysis. We move away from those kind of assumptions in our experimental

design and assume $w(z)$ to be a non-parametric flexible GP that is continuous. We use the best GP model from Chapter 3, which incorporates SNe, BAO, and CMB data to compare these different SNe telescope surveys.

There are three proposed future telescope surveys of SNe Type Ia: one ground based (LSST comprised of DEEP and MAIN) and two different space based missions (two space missions - JDEM and Wide-Field Infrared Survey Telescope (WFIRST), it would produce something like the simulated data we presented in Chapter 1) (Bernstein et al., 2009; Gehrels, 2010, 2011). The ground based surveys produce large numbers of SNe measurements more economically but the space based missions provide higher redshift observations with less observational uncertainty. We summarize these telescope surveys in Table 4.2 along with the details of the currently available SNe dataset, Union 2.

Table 4.2: Telescope Missions

	Current (Union2)	Ground (LSST)	Space 1 (JDEM)	Space 2
No. of SNe	557	10,000	1,500	2,300
z range	0.015-1.4	0-1.1	0.2 - 1.3	0-1.7
Mean z value	0.35	0.5	0.8	0.8
Error bars	0.08-1.08	0.12-0.16	0.12	0.13
Cost	n/a	n/a	1 bil	n/a

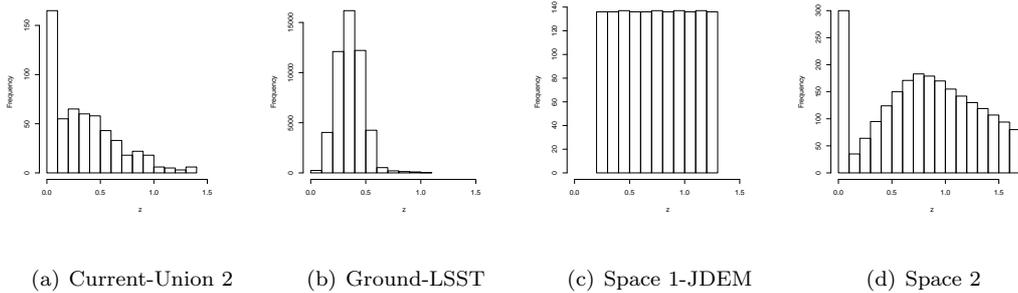


Figure 4.3: Redshift distributions for the three surveys

Underlying the different telescope surveys is a fundamental question about the optimal locations to collect data. This leads us to do an optimal design of experiments study. Optimality can be defined in many ways and we must first choose an optimality criteria for this experiment. Alphabetic optimality include criterion such as minimizing the variance or maximizing the information through a variety of methods: minimize the trace of the inverse information matrix (A-optimality) or minimize the determinant (D-optimality) or maximizes the minimum eigenvalue (E-optimality), or minimize the maximum entry in the diagonal of the hat matrix (G-optimality) or maximize the trace of the information matrix (T-optimality) and many others (DasGupta, 1995; Boyd and Vandenberghe, 2004). Some of the alphabetic criterion are to optimize the predictive variance like G-optimality, I-optimality and V-optimality. All of these methods are straight-forward when considering a linear model with fixed variance and non-hierarchical priors because they only require analysis of the design matrix (information matrix) (Chaloner and Verdinelli, 1995). The methods become increasingly more difficult or impossible if any one of these assumptions is changed.

We begin by stating the complexity of the problem at hand. The variance term is estimated. The favored model is a non-parametric infinite dimensional GP, in most cases its covariance matrix can be optimized by one of many methods (Etman, 1994). But in this problem, the GP is embedded in a non-linear transform which precludes such a simple optimization quantity as the covariance matrix. There can also be other issues because points tend to be chosen near the boundaries of the problem. The cosmology equation is non-linear and computationally expensive to perform with its double integration steps. Non-linear regression can have an optimality criterion where one minimizes the information matrix of the derivatives of the parameters (for D-optimality one would maximize the determinant of the Fisher information matrix or the parameters and E-optimality would maximize the minimum eigenvalue) (Berger, 2006; Dette et al., 2004). These methods all require a matrix

of the derivatives of the parameters (the Fisher matrix), which is complex because of the integrations inherent in the cosmology model. And this method does not extend well to a non-parametric model where there is an infinite domain for which the GP is defined. Suffice to say, no typical methods readily apply to the model of the Universe.

We thought of viewing this problem as multiple hypothesis testing on the $p = 55$ parameter space of the GP. The hypothesis testing for something as simple as $p = 3$ where $w(z) = \text{const}$ with point mass priors did not work (Toman, 1996). This is because it requires something similar to a RJMCMC algorithm with posteriors that did not converge in the MCMC algorithm. We thought about viewing this as a sequential design where we use the current Union 2 data, and set priors for Ω_m , $w(z)$, and the other parameters based on that analysis (Muller et al., 2004). But we would like to incorporate the current data with each of the future surveys in one cohesive model.

4.2.1 SNe Designed Experiments

We propose a computationally expensive alternative that adequately addresses the issues with other methods. We begin with the Union 2 dataset, which has $n = 557$ observations (z^o, μ^o) ; we include the one CMB data point and two BAO points. This analysis results in the fit for $w(z)$ shown in Figure 3.7(d); the process of fitting this model under the Bayesian framework and the prior settings are discussed in the preceding chapter. This results in posterior distributions for $\psi^o = (\Omega_m^o, \Delta^o, \sigma^{2o}, \rho^o, \kappa^{2o}, w(z^o))$ from the GP model (see Figure 3.7(d)). The posterior distributions for $w(z^o)$ is over a grid of fifty redshift values.

We want to compare the three possible telescope surveys to see which is the best at constraining the resulting $w^*(z)$. First, we must set a criteria to define what we mean by best constrains $w^*(z)$ or rather minimizes its uncertainty. We choose a criteria that to minimize the average variance of $w^*(z)$ over the grid of fifty points. We call the average variance M

(shown in Eq(4.1)) for each telescope survey and the $\min(M_{LSST}, M_{JDEM}, M_{Space2})$ should be the best of the three surveys at reducing the uncertainty associated with $w(z)$.

$$M = \frac{1}{\Delta z} \int \int \int V(w(z^*)|z^*, \mu^*, z^o, \mu^o, \psi^o, \psi^*) p(z^*, \mu^*, \psi^*|z^o, \mu^o, \psi^o) p(z^o, \mu^o, \psi^o) d\mu^* dz^* dz \quad (4.1)$$

where $\psi = (\Omega_m, \Delta, \sigma^2, \rho, \kappa^2)$, (z^o, μ^o) is the Union 2 observed data, and (z^*, μ^*) is simulated future data.

A set of values are drawn from the ψ distribution that are used to create new datasets. We create one hundred datasets in this fashion for each telescope survey. The distribution for the different z^* are shown in Figure 4.3 and the values for n^* are given in Table 4.2. We draw one hundred simulated datasets using z^* with n^* observations. The values of μ^* are based on $T(z^*, \psi^o|z^o, \mu^o)$ and a noise term $N(0, \epsilon^2)$, where ϵ is based on the values of the error bars in Table 4.2. Next we fit the GP model with parameters $w^*(z)$ and ψ^* to the one hundred new combined dataset $((z^o, z^*), (\mu^o, \mu^*))$ using the same priors for ψ^* as we did when we fit ψ^o with the original data (z^o, μ^o) .

As an overview, we begin with the real data (z^o, μ^o) and obtain $p(w(z^o), \psi^o|z^o, \mu^o)$; this produces one set of posteriors. The next steps, we perform one hundred times for each telescope survey. We draw simulated redshift locations z^* from the distribution, $p(z^*)$, in Figure 4.3. Next we obtain $\mu^*|z^*, w(z^o), \psi^o$ and draw error terms based on the error bars given in Table 4.2. Finally, we calculate $p(w^*(z), \psi^*|z^*, \mu^*, z^o, \mu^o)$. We have one hundred posteriors for $w(z)$ over a grid of fifty points. The variance for each of those 50 points is saved for each of the 100 different simulated SNe datasets. The mean of all of the variances is taken over all of the grid points and simulated sets, this gives an average overall variance. This produces one number for each of the surveys; we take the minimum to choose the best survey at shrinking the uncertainty associated with $w(z)$. We cannot say that the particular survey is the best because other factors can be considered through a loss function based on various costs. We leave that in the hands of the decision makers.

4.2.2 Results

The average variance of $w(z)$ for the simulated telescope surveys should be smaller than the average variance of the Union 2 data alone. This is the case for all three of the future telescope surveys. Union 2 has an initial average posterior, $E(V(w(z^o))) = 0.019$. For the one hundred simulated predictive datasets for each telescope mission, we obtain results as seen in Table 4.3. The JDEM mission has the most shrinkage for the average variance for $w(z)$ thus we would say this is the best mission. But none of the missions are statistically significantly different from one another in this analysis. And none are significantly different from the current Union 2 dataset, even though there are more SNe points with much tighter standard error bars (τ) for each of these three telescope missions.

Table 4.3: Telescope Mission Comparison

	$E(V(w(z^o)))$	90% PI
Union 2	0.019	n/a
Ground (LSST)	0.01480 ± 0.00341	(0.011264,0.022147)
Space 1 (JDEM)	0.01747 ± 0.00603	(0.011973,0.031743)
Space 2	0.01748 ± 0.00341	(0.012129,0.024950)

We believe there are two issues that arise when comparing the average variance for $w(z)$ for these different surveys. Some of the one hundred predictive dataset realizations for each telescope survey give a wider overall $V(w(z))$ than the Union2 set alone. Incompatible values of Ω_m and Δ between the Union2 set and the new predictive dataset could be causing this phenomenon. Thus there may be too much variation input in the creation of the predictive datasets. A second issue may also influence the lack of shrinkage of the PI bands for $w(z)$. In (Holsclaw et al., 2011b), we compare $n = 550$ data points to $n = 2300$ data points and the resulting $w(z)$ is quite similar for the two datasets even though one dataset is nearly four times larger. We think there may be an issue with saturation of the data; more data may not give much new information toward the estimation of $w(z)$.

Chapter 5

Conclusion

What conclusions can we make about the form of the dark energy EOS? Is $w(z)$ equal to negative one or with what certainty can we currently estimate $w(z)$? We may not currently be able to answer these questions fully. But we have made progress as to understanding the form of dark energy EOS and its relation to other parameters using various data sources.

In Chapter 1, we see the favored parametric representations of $w(z)$ are limited to one and two parameter models. The one parameter model, $w(z) = a$, has drawbacks in that it assumes a flat representation of $w(z)$. Ω_m and $w(z)$ are interdependent and different sets of these parameters produce equivalent solutions for the distance-redshift equation. We show a case (dataset μ_3) where the solution chosen by SNe data alone for these parameters is incorrect. Either a tighter prior on Ω_m resulting from BAO and CMB data is necessary or BAO and CMB data need to be directly included into the analysis. The two parameter models do not have these issues but they do make strict assumptions about the form of $w(z)$. Unfortunately, they cannot be extended to include more terms and thus more flexibility. There is a need for a more flexible model than what the literature has thus far proposed.

In Chapter 2, the non-parametric GP model and an approximation to that model using a basis of damped Hermite polynomials are presented. The approximation method has drawbacks similar to that of the two parameter models. We work with the inherent properties of the GP to propose a more stable and computationally efficient method that comes close to having all of the benefits of the approximation method. This is a novel approach to GP modeling that result in an integrated GP by using the properties of the stochastic process. The flexible GP model seems to out-perform the two parameter models by reducing the uncertainty in the estimate of $w(z)$. But $w(z)$ suffers from similar identifiability issues as some of the parametric models because Ω_m is not well constrained by the SNe data or its current prior.

At this point, we compare all of these models. We find that some of the model comparison methods being used in the literature are insufficient for this particular second derivative problem, in particular Bayes factors, AIC, and BIC. The DIC and PPL tests are used to compare the parametric and non-parametric models to a null hypothesis model where $w(z) = -1$. We conclude that there is currently not enough information in the SNe data to make a firm conclusion about the form of $w(z)$. At this time, we need to address the interdependencies of $w(z)$ and Ω_m that produce the multiple solution. And we need to examine what additional data is needed to better constrain $w(z) = -1$.

In Chapter 3, we introduce additional sources of information with different distance-redshift relationships than that of SNe. We use an informative prior on Ω_m in previous chapters resulting from BAO and CMB data sources. In this section, we trade the informative prior for a non-informative one and include the BAO and CMB data directly. Both BAO and CMB data have their own distance-redshift relation, which helps to uncouple $w(z)$ and Ω_m . We conclude that at least one additional data source should be included with any SNe analysis because it better informs Ω_m and helps decouple it from $w(z)$.

In Chapter 4, we desire to further constrain $w(z)$ than the current 550 SNe observations, 2 BAO points, and 1 CMB point can do. There are several astronomical surveys under consideration to collect additional data but they all cost a great deal of resources. We compare three of the SNe surveys being considered and the BAO survey in an experimental design. We set an optimality criterion for what it means to reduce uncertainty for the parameter space and perform the computations. We find that the ground survey provides the best reduction in uncertainty for $w(z)$ but that it is not statistically significant from other telescope surveys. We find that there may be a point at which more SNe data may not be providing further constraint for $w(z)$ because the data is sufficiently saturated.

Overall, the highlights of this work have been in successfully fitting a coherent non-parametric form to the dark energy EOS, $w(z)$. This alone is a sizable challenge; it allows for visualization of a time evolution form of dark energy (Holsclaw et al., 2010b,a). This requires statistical work on an inverse method for GPs when the curve of interest is a derivative process (Holsclaw et al., 2011a). This non-parametric GP model is further extended to include three types of data and better constrain Ω_m (Holsclaw et al., 2011b). Finally, we are able to draw on the three sources of data and the non-parametric model to perform a complex analysis of future astronomical surveys where the goal is to collect the best data to inform $w(z)$. Because of this work, we have a better understanding of the properties of the dark energy EOS and direction as to how to better understand it.

5.1 Future Work

There is so much to still discover about dark energy and its equation of state. Overall, science is still grappling with the issue of dark energies existence. If the suppositions are correct, we merely have been examining the form of its pressure and density relation. New data from the current sources will continue to be collected and need analysis to further

inform about this relation. This analysis in turn could drive the direction of future studies. We only investigated a handful of future SNe studies. As technology continues to improve and scientist learn even more there will be other astronomical studies to compare. The astronomers continue to reduce uncertainty with the measurements and light curve fitting process.

It would be interesting to look at the light curve fitting process from a statistical vantage point. It may be possible to incorporate the curve fitting process and modeling into one joint step. Statistically, it is better to incorporate all information into a unified model than to use summary statistics and uncertainty measures resulting from a black box process. Variables like color and shape from light curve fitting process could be incorporated as well into the analysis and their roles may be better understood. This line of study could possibly aid the observational astronomer and focusing their efforts at reducing uncertainty in specific areas. Any reduction in uncertainty should help to better explain the dark energy EOS.

Appendix A

GP Model: Details

A.1 Correlation Function Relationships

Assume we have the following relation between $y(z)$ and $w(z)$: $y(z) = \int_0^z w(u)du$ with $w(u) \sim GP(0, \Sigma_w = K(u, u'))$. We want to show that $y(z)$ is a Gaussian process with these specific properties: $y(z) \sim GP(0, \Sigma_y)$ with $\Sigma_y = \int_0^z \int_0^{z'} K(u, u')du'du$ and $\Sigma_{yw} = cov(y(z), w(u)) = \int_0^z K(u, u)du$ (we can obtain Σ_{wy} in a similar manner).

The integral of a GP is a GP because integration is approximately a linear operator (Sarkka, 2011; Rasmussen and Williams, 2006). We also can show that Y is a GP (or follows a multivariate Normal distribution) because $Y = \int_0^z w(u)du$. $w(u)$ is a GP by assumption and w is a continuous function of u . This is standard Riemann integration and can be approximated by the sum: $Y_n = \sum_{k=0}^{n-1} w(u_k)\Delta(u)$ where $Y_n \rightarrow Y$ and $n \rightarrow \infty$. $\sum_{k=0}^{n-1} w(u_k)$ is simply an n dimensional multivariate Normal. Therefore, each Y_n is Normal and this leads us to conclude that the process $y(z)$ is a GP.

Primarily, the focus is on zero mean GPs. But we want to show the properties of the a mean function that is non-zero. If $E(w(u)) = m(u)$ then $E(y(z)) = E(\int_0^z w(u)du) = \int_0^z E(w(u))du = \int_0^z m(u)du$ (we use Fubini's Theorem to exchange the expectation and the

limits of integration.)

Finally, we look at the properties of the covariance function of $y(z)$. We begin with the following definitions: $V(w(u)) = \text{cov}(w(u), w(u)) = K(u, u)$ and $\text{cov}(w(u), w(u')) = E(w(u)w(u')) = K(u, u')$ where $K(u, u')$ could be any correlation function; we mostly have been using the powered exponential correlation family. These definitions lead to the following:

$$\begin{aligned}
\Sigma_y &= \text{cov}(y(z), y(z')) \\
&= \text{cov}\left(\int_0^z w(u)du, \int_0^{z'} w(u')du'\right) \\
&= E\left(\int_0^z w(u)du \int_0^{z'} w(u')du'\right) \\
&= E\left(\int_0^z \int_0^{z'} w(u')w(u)du'du\right) \\
&= \int_0^z \int_0^{z'} E(w(u')w(u))du'du \quad \text{by Fubini's Thm} \\
&= \int_0^z \int_0^{z'} K(u, u')du'du
\end{aligned}$$

$$\begin{aligned}
\Sigma_{yw} &= \text{cov}(y(z), w(u')) \\
&= \text{cov}\left(\int_0^z w(u)du, w(u')\right) \\
&= E\left(\int_0^z w(u)du w(u')\right) \\
&= E\left(\int_0^z w(u')w(u)du\right) \\
&= \int_0^z E(w(u')w(u))du \quad \text{by Fubini's Thm} \\
&= \int_0^z K(u, u')du
\end{aligned}$$

We use Fubini's Theorem in these proofs but it requires that the function inside the integral have certain properties. The probability density function (pdf) satisfies these conditions because the expectation is always greater or equal to zero, as is the correlation

function $K(u, u')$. In general, with $f(z, z')$ being the pdf for the expectation we now show the properties of Fubini's Theorem in full detail:

$$\begin{aligned}
E(y(z)y(z')) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z, z')y(z)y(z')dz' dz \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z, z') \left(\int_0^z w(u)du \int_0^{z'} w(u')du' \right) dz' dz \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^z \int_0^{z'} f(z, z')w(u)w(u')du' dudz' dz \\
&= \int_0^z \int_0^{z'} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z, z')w(u)w(u')dz' dz du' du \\
&= \int_0^z \int_0^{z'} E(w(u), w(u'))du' du \\
&= \int_0^z \int_0^{z'} K(u, u')du' du
\end{aligned}$$

$f(z, z')w(u)w(u') \geq 0$ thus fulfilling Fubini's theorem and allowing the reordering of integration.

A.2 Integration of Correlation Functions

We use the Chebyshev-Gauss quadrature method for solving the single integral of the correlation function. $K(s, s') = \int_0^{s'} \frac{\rho^{|u-s|^\alpha}}{(1+u)} du$ is the integral equation used in the correlation matrix and it cannot be solved analytically. Additionally, there is no good approximation because the limits of integration do not go between 0 and infinity. Numerical methods are a must and Chebyshev-Gauss quadrature provides a good alternative to other forms of slower numerical integration (R uses Gauss-Kronrod quadrature).

First, we must change the limits of integration from $[0, s']$ to $[-1, 1]$ to be able to use this form of quadrature thus using the rule: $\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{(b-a)x}{2} + \frac{a+b}{2}\right)dx$. Chebyshev-Gauss quadrature uses Chebyshev polynomials of the first kind as its orthogonal polynomials, $\frac{1}{\sqrt{1-x^2}}$. In total, one integral is approximated as: $\int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}} dx \sim \sum_{i=1}^n \gamma_i f(x_i)$, $x_i = \cos\left(\frac{(2i-1)\pi}{2n}\right)$, and $\gamma_i = \frac{\pi}{n}$. We let: $u_i = \cos\left(\frac{(2i-1)\pi}{2n}\right)$, and $n = 100$ where the weights

are constant and fully specified. The single integral in equation (A.1) is equivalent to equation (A.2) which can be directly implemented into the GP algorithm.

$$K(s, s') = \int_0^{s'} \frac{\rho^{|u-s|^\alpha}}{(1+u)} du = \frac{s'}{2} \int_{-1}^1 \frac{\rho^{|\frac{s'u}{2} + \frac{s'}{2} - s|^\alpha}}{(1 + \frac{s'u}{2} + \frac{s'}{2})} \sqrt{1-u^2} \frac{1}{\sqrt{1-u^2}} du \quad (\text{A.1})$$

$$K(s, s') = \frac{s'}{2} \sum_{i=1}^n \gamma_i \frac{\sqrt{1-u_i^2}}{(1 + \frac{s'u_i}{2} + \frac{s'}{2})} \rho^{|\frac{s'u_i}{2} + \frac{s'}{2} - s|^\alpha} \quad (\text{A.2})$$

A.3 Localize Search Algorithm

There are numerical issues when using a GP model embedded in a highly non-linear transformation. Usually, a GP results in closed form posterior distributions that can easily be sampled from with Gibbs steps in an MCMC algorithm. But we have a GP model with n grid points that all must be sampled at every iteration through slow Metropolis steps in the MCMC algorithm. This step with n proposed points of the GP tends to have a high rejection level in the algorithm. To reduce the rejection level, the likelihood is transformed into an equivalent state that allows for smaller steps between the previous values. This can be thought of as a more localized search algorithm on the GP space.

A GP is defined by its mean and correlation functions: $w(z) \sim GP(\theta, \Sigma_{\rho, \kappa^2})$. The likelihood function resulting from the non-linear cosmology equations leads to the following posterior: $\sigma^2, \rho, \kappa^2 | \mu_i, \tau_i^2, z_i \propto L(z_i, \mu_i, \tau_i | w(z), \sigma^2) GP(w(z) | \rho, \kappa^2) \pi(\rho) \pi(\kappa^2) \pi(\sigma^2)$. Instead of the usual GP prior just described, we alter the form to allow for a more localized search. We let $w(z) \sim MVN(\theta, \Sigma)$ and $\Sigma^{-1/2}(w(z) - \theta) = w^o(z) \sim MVN(0, I)$. So the posterior becomes: $L(z_i, \mu_i, \tau_i | w^o(z), \rho, \kappa^2, \sigma^2) MVN(w^o(z); 0, I) \pi(\rho) \pi(\kappa^2) \pi(\sigma^2)$.

A.4 GP Algorithm for the Inverse Method

We thought the complete GP algorithm should appear somewhere to help clarify all of the complex steps. The algorithm contains all of the priors, the localized search method, and various methods of integration previously discussed.

1. Initialize all variables: $\rho = \rho_1$, $\kappa^2 = \kappa_1^2$, and $w^o(u) = w_{m,1}^o(u)$. $w(u)$ is a vector of m GP points and $y(s)$ is a GP with mh points. We run this algorithm $q = 1, \dots, Q$ times and the tuning parameters, $\delta_{1,2,3}$, need to be tuned until good mixing occurs. The variance parameters σ^2 and κ^2 must be greater than zero and ρ is in the range $(0, 1)$, any proposals that do not fit these criterion are rejected. All proposals are symmetric and do not need jumping functions.
2. Propose $\rho^* = \text{Unif}(\rho_q - \delta_1, \rho_q + \delta_1)$

(a) Compute the covariance matrix $K_{22\rho^*} = \rho^{*|u_j - u_i|^\alpha}$

(b) Compute the Cholesky decomposition for $K_{22\rho^*} = U_{\rho^*}' U_{\rho^*}$

(c) Compute the special $K_{12\rho^*} = \int_0^{s'} \frac{\rho^{*|u-s|^\alpha}}{1+u} du$

(d) We want $y_{\rho^*}(s) = \theta \ln(1+s) + [\kappa_{q-1}^2 K_{12*}] [\kappa_{q-1}^2 K_{22*}^{-1}] (w_{\rho^*}(u) - \theta)$

where: $w_{\rho^*}(u) = [\kappa_{q-1} U_{\rho^*}'] w_{m,q-1}^o + \theta$

$$y_{\rho^*}(s) = \theta \ln(1+s) + [\kappa_{q-1}^2 K_{12*}] [\kappa_{q-1}^2 K_{22*}^{-1}] ((\kappa_{q-1} U_{\rho^*}' w_{m,q-1}^o + \theta) - \theta)$$

$$= \theta \ln(1+s) + \kappa_{q-1} K_{12*} [(U_{\rho^*}' U_{\rho^*})^{-1} U_{\rho^*}'] w_{m,q-1}^o$$

$$= \theta \ln(1+s) + \kappa_{q-1} K_{12*} [U_{\rho^*}^{-1}] w_{m,q-1}^o$$

(e) $L(z_i, \mu_i, \tau_i | w_{\rho^*}, \sigma_{q-1}^2) = e^{-\frac{1}{2} \sum \left(\frac{\mu_i - T(z_i, w_{\rho^*}(u))}{\tau_i \sigma_i} \right)^2}$ where the definite integrations

in $T(z_i, w_{\rho^*}(u))$ are done numerically through summations of the trapezoid algorithm.

(f) If we accept $\alpha_{MH} = \frac{L_{\rho^*} \pi(\rho^*)}{L_{\rho_{q-1}} \pi(\rho_{q-1})}$ then we let $\rho_q = \rho^*$

3. Draw $\kappa^{2*} = Unif(\kappa_{q-1}^2 - \delta_2, \kappa_{q-1}^2 + \delta_2)$
 - (a) Compute $y_{\kappa^{2*}}(s) = \theta \ln(1 + s) + \kappa^* K_{12\rho_q} [U_{\rho_{q-1}}^{-1}] w_{m,q-1}^o$
 - (b) $L(z_i, \mu_i, \tau_i | w_{\kappa^{2*}}, \sigma_{q-1}^2) = e^{-\frac{1}{2} \sum \left(\frac{\mu_i - T(z_i, w_{\kappa^{2*}}(u))}{\tau_i \sigma_i} \right)^2}$ where the definite integrations in $T(z_i, w_{\kappa^{2*}}(u))$ are done numerically through summations of the trapezoid algorithm
 - (c) If we accept $\alpha_{MH} = \frac{L_{\kappa^{2*}} \pi(\kappa^{2*})}{L_{\kappa_{q-1}^2} \pi(\kappa_{q-1}^2)}$ then we let $\kappa_q^2 = \kappa^{2*}$
4. Propose a non-standard w_m^* as the GP. Start by drawing a proposal for $w^{o*} \sim MVN(w_{q-1}^o, \delta_3 I_{m \times m})$
 - (a) Compute $y^*(s) = \theta \ln(1 + s) + \kappa_q K_{12q} [U_q^{-1}] w_m^{o*}$,
 - (b) $L_{z_i, \mu_i, \tau_i | w_{new}^*(u), \sigma_{q-1}^2} = e^{-\frac{1}{2} \sum \frac{\mu_i - T(z_i, w_{new}^*(u))}{\tau_i \sigma_i}}^2$
 - (c) If we accept $\alpha_{MH} = \frac{L_{w_{new}^*(u)} MVN(w_m^{o*} | 0, I)}{L_{w_{q-1}^o} MVN(w_{m,q-1}^o | 0, I)}$ then $w_{m,q}^o(u) = w_m^{o*}(u)$ and the Gaussian process realization is $w_{m,q}(u) = w_m^*(u)$
5. $\sigma_q^2 | \dots \sim IG\left(\frac{n}{2}, \frac{1}{2} \sum \left(\frac{\mu_i - T(z|\dots)}{\tau_i} \right)^2\right)$
6. Repeat steps 2-6, Q times and rerun the entire algorithm as needed after resetting the tuning parameters.

A.5 MCMC Issues

MCMC issues continually arise for this cosmology application as the complexity of the non-linear transform make parameter uncertainty estimation difficult. There is the possibility of closed form posterior distributions for two specific parameters: σ^2 and Δ . But Ω_m and the parameters that comprise $w(z)$ do not have closed form posterior distributions, thus making sampling from them troublesome. Many additional issues arise for the typical MCMC sampling procedures. In the simple cases, where $w(z)$ has a parametric form, the

standard non-linear regression methodology works. But here the parameters tend to be correlated to one another, so we adopt joint proposal distributions for the parameters. The convergence and stability of the posterior chains are greatly improved with this method.

In standard methods, the hyper-parameters in the correlation function of the GP typically do not have closed form posterior distributions but the GP itself does. In this cosmology application, the GP does not have a closed form posterior, leading to slow sampling of the posterior chain. We implement a localized search algorithm to aid with the mixing and convergence of the posterior chains. The acceptance rates for the proposals of the GP points are still lower than usual, closer to 10% than 30%. All GP points are sampled in a joint step, the two correlation parameters are sampled in their own joint step, as are Δ and Ω_m . This algorithm poses many issues with computational efficiency that we discuss in Chapter 2. Additionally, the grid points of the GP algorithm also control the speed and accuracy of the results of the integration steps. We do not use an evenly spaced grid over the GP domain. We have more grid points near zero on the z axis because the integrations are performed from zero to z_i for every point, $i = 1 \dots n$.

Issues arise when sampling from the posteriors without closed form. Hypothesis testing, for whether $w(z)$ is equal to negative one, poses the same issues with mixing and convergence. We assume point mass priors for the parameters of $w(z)$ in this method. It requires a type of reversible jump MCMC (RJMCMC) algorithm. But the posterior distributions we obtain are linked to the proposal distribution. The percentage of time the posterior stays in the point mass is dependent on the proposals tuning parameter. In most cases, convergence is not properly obtained, so we abandon this method altogether and do not display results. The literature on Bayesian hypothesis testing using point mass priors, seems to only address cases where closed form posteriors are available and proposal distributions are unnecessary. Instead, we pursue model comparison methods and use a null hypothesis model for hypothesis testing.

A.6 Correlation Parameter

There are several ways of representing the correlation parameter in the exponential family: $K(u, u') = \rho^{|u-u'|^\alpha}$. We have chosen this parameterization with ρ but to better understand the properties of this parameter. But it may be worthwhile to look at other common parameterizations like: $\rho = \exp(-\lambda) = \exp(-1/d)$. We examine in Figure A.1 two possible priors for ρ and the induced priors on λ and d . The first is a flat prior for ρ , which corresponds to the $IG(1,1)$, d then has a mode of one half and undefined mean and variance. The second is an informative prior for ρ of $(\text{Beta}(6,1))$ that emphasizes smoothness of the GP, which corresponds to $IG(1,6)$ prior on d . We choose this second prior because the exponential correlation function is quite jagged and we want to impose some smoothness through the hyper-parameter prior.

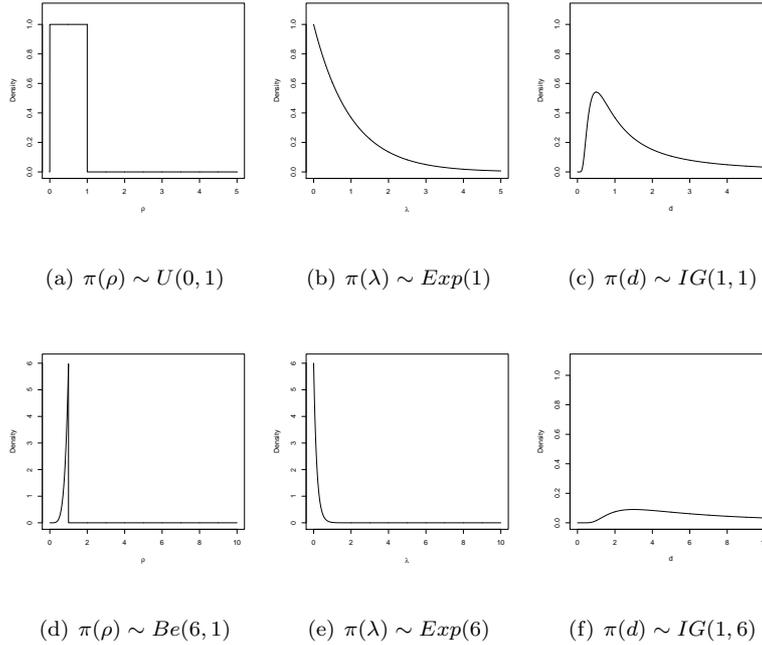


Figure A.1: Examining different parameterizations of the correlation parameter

The parameter d is of interest because it best explains the range of correlation function; the rest of the discussion is in terms of d and not ρ . It makes sense to discuss the modes of these prior distributions (except for the $U(0,1)$) because they are skew and the mean is not defined in all cases (the IG distribution used here does not have a defined mean or variance). The first option of prior for d has a mode of one half and the second has a mode of three.

We use the definition of practical range to be when the correlation is equal to 0.05 and this range depends on both d and α (Diggle and Ribeiro, 2007). The redshift data ranges from 0 to about 1.7; some may want the practical range (correlation lengths) within these values. For the exponential correlation function ($\alpha = 1$) the practical range is approximately $3d$, for the Gaussian correlation function ($\alpha = 2$) it is $\sqrt{3}d$ and for the Matérn($\nu = 1.5$) it is $4.75d$. The prior $d \sim IG(1, 1)$ has a mode of 0.5 which has a practical range very near 1.7 but the resulting GP on $w(z)$ is far too rough. We use the second prior and the correlation is 0.72 at redshift 1.7; this prior implies that the data on the whole observed spectrum is highly related.

Two things we assume from the physics of the problem: first $w(z)$ is completely continuous and second it is rather smooth (no sudden changes). The GP is assumed to be continuous (because of the choice of correlation function) and fully defined at every point on the redshift range. The smoothness of the GP must also be inferred in the choice of correlation function (or in other words the value of the α parameter for the exponential family or ν for the Matérn) and the prior on d . Even though, the distributional form of the prior on d is identical for the exponential and Gaussian correlation functions, its practical range is different. When we have a Gaussian correlation, we continue to use the same prior but now the practical range is smaller, about $\sqrt{3}d$ instead of $3d$. The Gaussian correlation is infinite differentiability and very smooth, so relaxing the practical range makes sense.

We show the difference between two priors for ρ and the resulting $w(z)$ fits in Figure A.2. The Uniform prior on ρ produces a fit for $w(z)$ that is jagged and does not hold to the physics of the problem. This results in much wider bands on $w(z)$ because the initial assumptions of smoothness are missing. In Figure A.3, we plot the prior and posterior distribution for ρ for these two different prior assumptions. The prior greatly effects the posterior. The other thing to note is when we have a flat prior on ρ there is still a tendency toward high values for ρ near one. It is not just our informative prior that tends towards high correlation lengths. Overall, we use the informative prior with additional smoothness assumptions and the many simulated data examples show that it performs quite well.

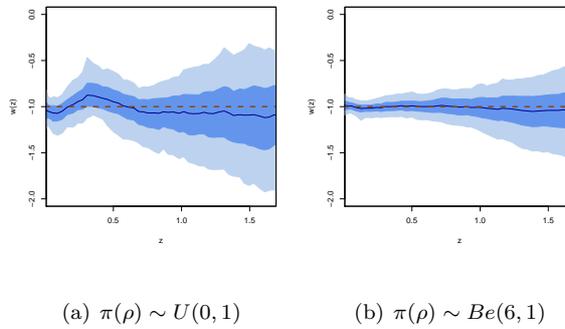


Figure A.2: Model 4 fits of $w(z)$ for dataset μ_1 for different priors on ρ

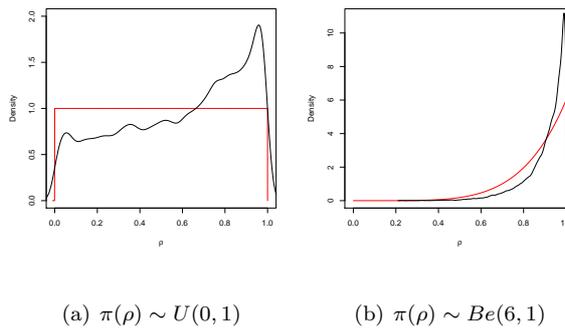


Figure A.3: The black line is the posterior density of ρ and red line is the prior density

Bibliography

- Aldering, G. (2005), “Exploring Dark Energy with SNAP,” *New Astron.Rev.*, **49**.
- Amanullah, R., Lidman, C., Rubin, D., and Aldering, G. *et al.* (2010), “Spectra and Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation,” *Astrophys. J.*, **716**.
- Astier, P. (2000), “Can Luminosity Distance Measurements Probe the Equation of State of Dark Energy?” *Physics Letters B*, **500**, 8–15.
- Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, New York: Chapman and Hall.
- Banfield, J. and Raftery, A. (1993), “Model-Based Gaussian and Non-Gaussian Clustering,” *Biometrics*, **49**.
- Bennett, C., Bay, M., Halpern, M., and Hinshaw, G. *et al.* (2003a), “The Microwave Anisotropy Probe (MAP) Mission,” *ApJ*, **583**.
- Bennett, C., Halpern, M., Hinshaw, G., and Jarosik, N. *et al.* (2003b), “First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results,” *ApJS*, **148**.
- Berger, J. (2006), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer.

- Bernstein, J., Cinabro, D., Kessler, R., and Kuhlman, S. (2009), “LSST Science Book,”
ArXiv e-prints - arXiv:0912.0201v1.
- Bond, J., Efstathiou, G., and Tegmark, M. (1997), “Forecasting Cosmic Parameter Errors
from Microwave Background Anisotropy Experiments,” *Not. Roy. Astron. Soc.*, **291**, L33–
L41.
- Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, New York: Cambridge Uni-
versity Press.
- Chaloner, K. and Verdinelli, I. (1995), “Bayesian Experimental Design: A Review,”
Statist.Sci, **10**.
- Chevallier, M. and Polarski, D. (2001), “Accelerating Universes with Scaling Dark Matter,”
Int.J.Mod.Phys.D., **10**.
- Cooray, A. and Huterer, D. (1999), “Gravitational Lensing as a Probe of Quintessence,”
Astrophys. J., **513**.
- Corasaniti, P. and Melchiorri, A. (2008), “Testing Cosmology with Cosmic Sound Waves,”
Phys.Rev.D, **77**.
- Crittenden, R. and Pogosian, L. (2005), “Investigating Dark Energy Experiments with Prin-
cipal Components,” *arXiv:astro-ph/0510293v1*.
- DasGupta, A. (1995), “Review of Optimal Bayes Designs,” *Technical Report - Purdue Uni-
versity*.
- Davis, T., Mortsell, E., and Sollerman, J. *et al.* (2007), “Scrutinizing Exotic Cosmological
Models Using ESSENCE Supernova Data Combined with Other Cosmological Probes,”
Astrophysics Journal, **666**, 716.

- de Sitter, W. (1934), “On Distance, Magnitude, and Related Quantities in an Expanding Universe,” *Bulletin of the Astronomical Institutes of the Netherlands*, **7**, 205.
- Dette, H., Melas, V., and Pepelyshev, A. (2004), “Optimal Designs for a Class of Nonlinear Regression Models,” *Bulletin of the Astronomical Institutes of the Netherlands*, **32**, 2142.
- Diggle, P. and Ribeiro, P. (2007), *Model-based Geostatistics*, New York: Springer.
- Dodelson, S. (2003), *Modern Cosmology*, Academic Press.
- Elgaroy, O. and Multamaki, T. (2007), “On Using the CMB Shift Parameter in Tests of Models of Dark Energy,” *Aston. and Astrophys.*, **471**.
- Escamilla-Rivera, C., Lazkoz, R., Salzano, V., and Sendra, I. (2011), “Tension Between SN and BAO: Current Status and Future Forecasts,” *ArXiv e-prints*.
- Etman, L. (1994), “Design and Analysis of Computer Experiments: The Method of Sacks et al,” *Engineering Mechanics Report WFW 94.098*, Eindhoven University of Technology.
- Feng, C., Wang, B., Abdalla, E., and Su, R. (2008), “Observational Constraints on the Dark Energy and Dark Matter Mutual Coupling,” *Physics Letters B*, **665**, 111119.
- Filippenko, A. (1997), “Optical Spectra of Supernovae,” *Annual Review of Astronomy and Astrophysics*, **35**, 309–355.
- Fraley, C. and Raftery, A. (2006), “MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering,” *Technical Report University of Washington*, **504**.
- Gamerman, D. and Lopes, H. (2006), *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, New York: Chapman and Hall.
- Gehrels, N. (2010), “The Joint Dark Energy Mission (JDEM) Omega,” *ArXiv e-prints*.
- (2011), “Wide-Field Infrared Survey Telescope (WFIRST),” *NASA*.

- Gelfand, A. and Ghosh, S. (1998), “Model Choice: A Minimum Posterior Predictive Loss Approach,” *Biometrika*, **85**, 1–11.
- Gelman, A., Carlin, B., Stern, H., and Rubin, D. (2004), *Bayesian Data Analysis*, New York: Chapman and Hall.
- Genovese, C., Freeman, P., Wasserman, L., Nichol, R., and Miller, C. (2009), “Inference for the Dark Energy Equation of State Using Type Ia Supernova Data,” *Annals of Applied Statistics*, **3**, 144–178.
- Guy, J., Astier, P., Nobili, S., Regnault, N., and Pain, R. (2005), “SALT: A Spectral Adaptive Light Curve Template for Type Ia Supernovae,” *Astrophysics*.
- Hicken, M., Wood-Vasey, M., Blondin, S., and Challis, P. *et al.* (2009), “Improved Dark Energy Constraints from ~ 100 New CfA Supernova Type Ia Light Curves,” *Astrophys.J.*, **700**.
- Higdon, D., Lee, H., and Holloman, C. (2003), “Markov chain Monte Carlo-Based Approaches for Inference in Computationally Intensive Inverse Problems,” in *Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting*, ed. Bernardo, J. *et al.*, Oxford University Press.
- Hojjati, A. and Pogosian, L. (2010), “Detecting Features in the Dark Energy Equation of State: a Wavelet Approach,” *Journal of Cosmology and Astroparticle Physics*.
- Holsclaw, T., Alam, U., Sansó, B., Lee, H., Heitmann, K., Habib, S., and Higdon, D. (2010a), “Nonparametric Dark Energy Reconstruction from Supernova Data,” *Physical Review Letters*, **105**.
- (2010b), “Nonparametric Reconstruction of the Dark Energy Equation of State,” *Physical Review D*, **82**.

- (2011a), “Gaussian Process Modeling of Derivative Curves,” *under review*.
 - (2011b), “Nonparametric Reconstruction of the Dark Energy Equation of State from Diverse Data Sets,” *under review*.
- Huterer, D. and Starkman, G. (2003), “Parameterizations of Dark-Energy Properties: a Principal-Component Approach,” *Phys. Rev. Lett.*, **90**.
- Huterer, D. and Turner, M. (1999), “Prospects for Probing the Dark Energy via Supernova Distance Measurements,” *Phys. Rev. D*, **60**.
- (2001), “Probing Dark Energy: Methods and Strategies,” *Phys. Rev. D*, **64**.
- Kaipio, J. and Somersalo, E. (2004), *Statistical and Computational Inverse Problems*, Springer-Verlag.
- Komatsu, E., Smith, K., Dunkley, C., and Bennett, B. *et al.* (2010), “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation,” *Astrophysical Journal Supplement Series*.
- Kowalski, M. and Rubin, D. *et al.* (2008), “Improved Cosmological Constraints from New, Old and Combined Supernova Datasets,” *Astrophysics*, **686**, 749–778.
- Krauss, L., Jones-Smith, K., and Huterer, D. (2007), “Dark Energy, a Cosmological Constant, and Type Ia Supernovae,” *New Journal of Physics*, **9**, 141.
- Leibundgut, B. (2001), “Cosmological Implications from Observations of Type Ia Supernovae,” *Annual Review of Astronomy and Astrophysics*, **39**, 67–98.
- (2004), “Are Type Ia Supernovae Standard Candles?” *Astrophysics and Space Science*, **290**.
- Linder, E. (2003), “Exploring the Expansion History of the Universe,” *PhysRevLett*, **90**.

- (2006), “Biased Cosmology: Pivots, Parameters and Figures of Merit,” *Astroparticle Physics*, **26**, 102–110.
- (2007), “The Mirage of $w=-1$,” *Astrophysics*.
- Maor, I., Brustein, R., and Steinhardt, P. (2001), “Limitations in Using Luminosity Distance to Determine the Equation of State of the Universe,” *Phys.Rev.Lett.*, **87**.
- March, M., Starkman, G., Trotta, R., and Vaudrevange, P. (2010), “Should We Doubt the Cosmological Constant?” *Mon.Not.R.Astron.Soc.*
- Muller, P., Sanso, B., and De Iorio, M. (2004), “Optimal Bayesian Design by Inhomogeneous Markov Chain Simulation,” *Journal of the American Statistical Association*.
- Neal, R. (1997), “Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification,” *Technical Report No. 9702, Department of Statistics, University of Toronto*.
- O’Hagan, A. (1991), “Bayes-Hermite Quadrature,” *Journal of Statistical Planning and Inference*, **29**, 245–260.
- Percival, W., Reid, B., Eisenstein, D., and Bahcall, N. *et al.* (2010), “Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Data Release 7 Galaxy Sample,” *Monthly Notices of the Royal Astronomical Society*, **401**, 2148–2168.
- Perlmutter, S., Aldering, G., Deustua, S., and Fabbro, S. *et al.* (1997), “Cosmology from Type Ia Supernovae,” *Bulletin of the American Astronomical Society*, **29**, 1351.
- Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R., and Nugent, P. *et al.* (1999), “Measurements of Omega and Lambda for 42 High-Redshift Supernovae,” *Astrophysics Journal*, **517**, 565–586.

- Pskovskii, I. (1977), “Light Curves, Color Curves, and Expansion Velocity of Type I Supernovae as Functions of the Rate of Brightness Decline,” *Soviet Astronomy*, **21**, 675.
- Rasmussen, C. and Williams, K. (2006), *Gaussian Processes for Machine Learning*, MIT Press.
- Riess, A., Filippenko, A., Challis, P., and Clocchiattia, A. *et al.* (1998), “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant,” *The Astronomical Journal*, **116**, 1009–1038.
- Riess, A., L.G., S., Casertano, S., and Ferguson, H. *et al.* (2007), “New Hubble Space Telescope Discoveries of Type Ia Supernovae at $z > 1$: Narrowing Constraints on the Early Behavior of Dark Energy,” *The Astrophysics Journal*, **659**, 98–121.
- Riess, A., Press, W., and Kirshner, R. (1996a), “Is the Dust Obscuring Supernovae in Distant Galaxies the Same as Dust in the Milky Way?” *The Astronomical Journal*, **473**, 588–594.
- (1996b), “A Precise Distance Indicator: Type Ia Supernova Multicolor Light Curve Shapes,” *The Astrophysical Journal*, **473**, 88–109.
- Sahni, V. and Starobinsky, A. (2006), “Reconstructing Dark Energy,” *Int.J.Mod.Phys.D*, **15**, 2105.
- Saini, T., Raychaudhury, S., Sahni, V., and Starobinsky, A. (2000), “Reconstructing the Cosmic Equation of State from Supernova Distances,” *Phys.Rev.Lett*, **85**, 1162–1165.
- Sarkka, S. (2011), “Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression,” *Proceedings of International Conference on Artificial Neural Networks (ICANN)*.

- Schlegel, D., Bebek, C., Heetderks, H., and Ho, S. *et al.* (2009a), “BigBOSS: The Ground-Based Stage IV Dark Energy Experiment,” *ArXiv e-prints*.
- Schlegel, D., White, M., and Eisenstein, D. (2009b), “The Baryon Oscillation Spectroscopic Survey: Precision Measurements of the Absolute Cosmic Distance Scale,” *ArXiv e-prints*.
- Schwarz, G. (1978), “Estimating the Dimension of a Model,” *Annals of Statistics*, 6, 461–464.
- Seo, H. and Eisenstein, D. (2007), “Improved Forecasts for the Baryon Acoustic Oscillations and Cosmological Distance Scale,” *Astrophys. J.*, **665**.
- Simpson, F. and Bridle, S. (2006), “Redshift Sensitivities of Dark Energy Surveys,” *Phys.Rev.D*, **73**.
- Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002), “Bayesian Measures of Model Complexity and Fit,” *Journal of the Royal Statistical Society B*, **64**, 583–639.
- Stein, M. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer.
- Steinberg, D. and Burstzryn, D. (2004), “Data Analytic Tools for Understanding Random Field Regression Models,” *Technometrics*, **46**.
- Szydlowski, M. and Wlodzimierz, G. (2006), “Which Cosmological Model - with Dark Energy or Modified FRW Dynamics?” *Physics Letters B*, **633**, 427–432.
- Tan, P., Steinback, M., and Kumar, V. (2006), *Introduction to Data Mining*, Boston: Addison-Wesley.
- Toman, B. (1996), “Bayesian Experimental Design for Multiple Hypothesis Testing,” *JASA*, **91**.

- Wang, Y. and Mukherjee, P. (2007), “Observational Constraints on Dark Energy and Cosmic Curvature,” *Phys.Rev.D*, **76**.
- Way, M., Foster, P., Gazis, P., and Srivastava, A. (2009), “New Approaches to Photometric Redshift Prediction via Gaussian Process Regression in the Sloan Digital Sky Survey,” *Astrophys.*, **706**.
- Weller, J. and Albrecht, A. (2001), “Opportunities for Future Supernova Studies of Cosmic Acceleration,” *Phys.Rev.Lett.*, **86**.
- (2002), “Future Supernovae Observations as a Probe of Dark Energy,” *Phys.Rev.D*, **65**.
- Wood-Vasey, W., Friedman, A., Bloom, J., and Hicken, M. *et al.* (2008), “Type Ia Supernovae are Good Standard Candles in the Near Infrared: Evidence from PAIRITEL,” *ApJ*, **689**.
- Wood-Vasey, W., Miknaitis, G., Stubbs, C., and Jha, S. *et al.* (2007), “Observational Constraints on the Nature of Dark Energy: First Cosmological Results from the Essence Supernova Survey,” *Astrophysics Journal*, **666**.
- Yaglom, A. (1962), *An Introduction to the Theory of Stationary Random Functions*, New York: Dover Publications, Inc.